MANUAL

(CytoSpec v. 2.00.xx)

Biomedical Applications of Vibrational Spectroscopy - Hyperspectral Imaging - Chemometrics

[Peter Lasch | contact: info@cytospec.com

CytoSpec - AN APPLICATION FOR HYPERSPECTRAL IMAGING

- Introduction/Preface
- Installation, System Requirements
- Main Window - Basic Concepts
- Hot Keys
- Data Organization
- Working with Spectra - Basic Concepts
- How to Produce Hyperspectral Maps
- The Log-File
- Ordering the Program, License, Disclaimer
- CytoSpec 64-bit
- What's new in CytoSpec 2.00.01?
- Download a free demo version (limited functionality)

File Pulldown Menu

- Load
- Save
- Save Matlab
- Import ASCII
- Import Binary
- Export
- Delete
- Clear
- Plot
- Customize
- Batch Multiple Files
- Exit

Spectral Preprocessing

- Calculation of Derivative Spectra
- Normalization (Vector, Offset)
- Cut
- Interpolate
- Smooth
- ABS <--> TR Conversion
- Subtraction
- Dispersion Correction
- Quality Test
- Baseline Correction
- Water Vapor Compensation
- Noise Correction
- Cosmic Spike Removal
- Batch Preprocessing

Spatial Preprocessing

- Crop
- Interpolate/Binning
- Filter Images
- 3D-FSD

Univariate Imaging

- Chemical Imaging
- Chemical Movie
- Frequency Imaging

Multivariate Imaging

- HCA Imaging
- KMC Imaging
- FCM Cluster Imaging
- PCA Imaging
- VCA Imaging
- n-findr Imaging
- ANN Imaging
- Synthon Imaging
- Imaging with Distance Values

Tools

- Display Spectra
- Set Display Limits
- Grid On/Off
- Set Colors
- Capture
- Export Maps
- Map Statistics
- Display Large Maps
- Define ROI
- Display Colorbar
- Swap Data Blocks
- Rotate
- Flip

File Information

- Show History
- Show Instrument Parame
- Show measurement Para
- Show Additional Paramet
- Edit Parameters
- About
- Using the Help Function

Preface

The CytoSpec program was developed in the year 2000 by Dr. Peter Lasch who worked as a postdoc with Prof. Max Diem in the Laboratory for Spectral Diagnosis at Hunter College (CUNY, New York), now Northeastern University (Boston). Starting from the first version in 2000 the software has been continuously improved and updated. CytoSpec is now known as a specialized software package for vibrational hyperspectral imaging that supports a large number of different tasks for data import, spectral and spatial preprocessing and uni- and multivariate image segmentation. The program is widely used in a scientific environment and was the workhorse in many scientific studies. The CytoSpec software has been continuously developed over the past years and supports now most of the actual hardware platforms, including 64-bit environments, see 64-bit version of CytoSpec for details.

How to obtain the program:

The full program (commercial version) can be ordered from CytoSpec. Please e-mail your request to CytoSpec (e-mail: order@cytospec.com). You can also obtain a free demo version (limited functionality) of CytoSpec (e-mail:service@cytospec.com).

To obtain the most recent version of the CytoSpec online help go to the ➡ CytoSpec web pages.

Introduction to the program / short description:

The software package available from CytoSpec is a program designed specifically for the analysis of vibrational spectroscopic (IR and Raman) imaging data sets. The major innovation of the CytoSpec software over competing software packages is its structure that manipulates hyperspectral data cubes, rather than individual spectra. Thus, all operations carried out on the data affect every spectrum, with the number of spectra in the data set limited only by the available memory.

The CytoSpec software is a stand-alone, comprehensive package that operates under Windows XP, Windows Vista and Windows 7. Hyperspectral data sets, in the format defined by the manufacturer of the Raman, or infrared, microspectrometer used by the researcher, are imported, converted to and stored in a data matrix format specific to the CytoSpec program.

The software permits the standard spectral manipulations customarily found in single spectra analysis software, such as expansion, smoothing, scaling, normalization, etc. Due to the fact that data sets often contain hundreds or thousands of spectra, a number of statistical approaches to the data are built in the software. These can be classified as uni- and multivariate statistical methods. Univariate methods of analysis, included in the software, consist of various mapping displays of hyperspectral data. In these, the user may select band intensities, integrated intensities, frequencies, intensity ratios, etc., to construct false color maps of the spectral data, which may be considered to be slices through the hyperspectral data cube.

The multivariate methods of data analysis create spectral correlations and maps by including not just one intensity or frequency point of a spectrum, but by utilizing the entire spectral information. These methods include principal component analysis (PCA), unsupervised methods of cluster analysis, and endmember selection methods. The software is configured to permit output of the spectral data in the format used by other software packages, such as the NeuroDeveloper (Synthon), an artificial neural network simulator.

SYSTEM REQUIREMENTS:

Software:

- Windows XP, Windows Vista or Windows 7. Windows 98 or ME are not recommended!
- Windows 64-bit versions are supported, but note that the compiled version of CytoSpec is still a 32-bit application. The memory advantage of the 64-bit OS is therefore not fully used (see also CytoSpec 64-bit).
- Software to display these help files, that is a web browser such as Firefox, Internet Explorer or Opera

Hardware:

- approx. 100 MB of free disk space for example files
- at least 1024 MB of RAM
- i386 based CPUs (Intel/AMD)
- at least an 8 bpp 960x720 graphic display. Supported display modes:
  960 x 720 (minimum)
  1024 x 768
  1152 x 864
  1280 x 1024
  1600 x 1200
  and most wide screen resolutions.
- to run CytoSpec 64-bit a full 64-bit hard- and software environment is required. Please refer to the CytoSpec 64-bit documentation for details.

INSTALLATION (32-bit)

Commercial version: Once you have received the CytoSpec CD or downloaded the installation archive from CytoSpec's web server, you are ready to install the program. During the install procedure CytoSpec's files will be automatically extracted from the cabinet and all necessary steps for starting CytoSpec will be taken.

- If you have downloaded CytoSpec: unzip the installation archive into a directory of choice. In this directory locate the file '*setup.exe*'
- CD-Version: open the Explorer and locate the CytoSpec setup file on the CD you have received. The file is named '*setup.exe*'
- To start the installation process just double click on setup.exe. The installation program will be started automatically.
- Determine the place on your hard disk where you want CytoSpec to be installed. Normally the proposed C:\program files\CytoSpec\CytoSpec will do fine.
- Then simply follow the instructions of the installation program.
- When setup is complete you can find a new program group and a CytoSpec shortcut on the desktop.
- Start the program by double clicking on the CytoSpec icon.

Demo version: Download the installation package from CytoSpec's web server. **IMPORTANT**: In order to run CytoSpec, an additional license file ('keygen.gen') must be copied into CytoSpec's main directory. This individual license key is NOT included in the installation package and can be obtained only by e-mailing a request to order@cytospec.com

- Unzip the installation archive into a directory of choice. In this directory locate the file '*setup.exe*'
- To start the installation process just double click on setup.exe. The installation program will be started automatically.
- Determine the place on your hard disk where you want CytoSpec to be installed. Normally the proposed C:\program files\CytoSpec\CytoSpec will do fine.
- Then simply follow the instructions of the installation program.
- When setup is complete you can find a new program group and a CytoSpec shortcut on the desktop.
- You can copy now the license key file into CytoSpec's main directory (e.g. c:\program files\CytoSpec\CytoSpec)
- Start the program by double clicking on the CytoSpec icon.
- Note that the temporary license for CytoSpec demo will expire after 90 days.

Patching CytoSpec: To manually update (patch) an older CytoSpec version you can simply overwrite the file 'ftir.exe' by a newer one.

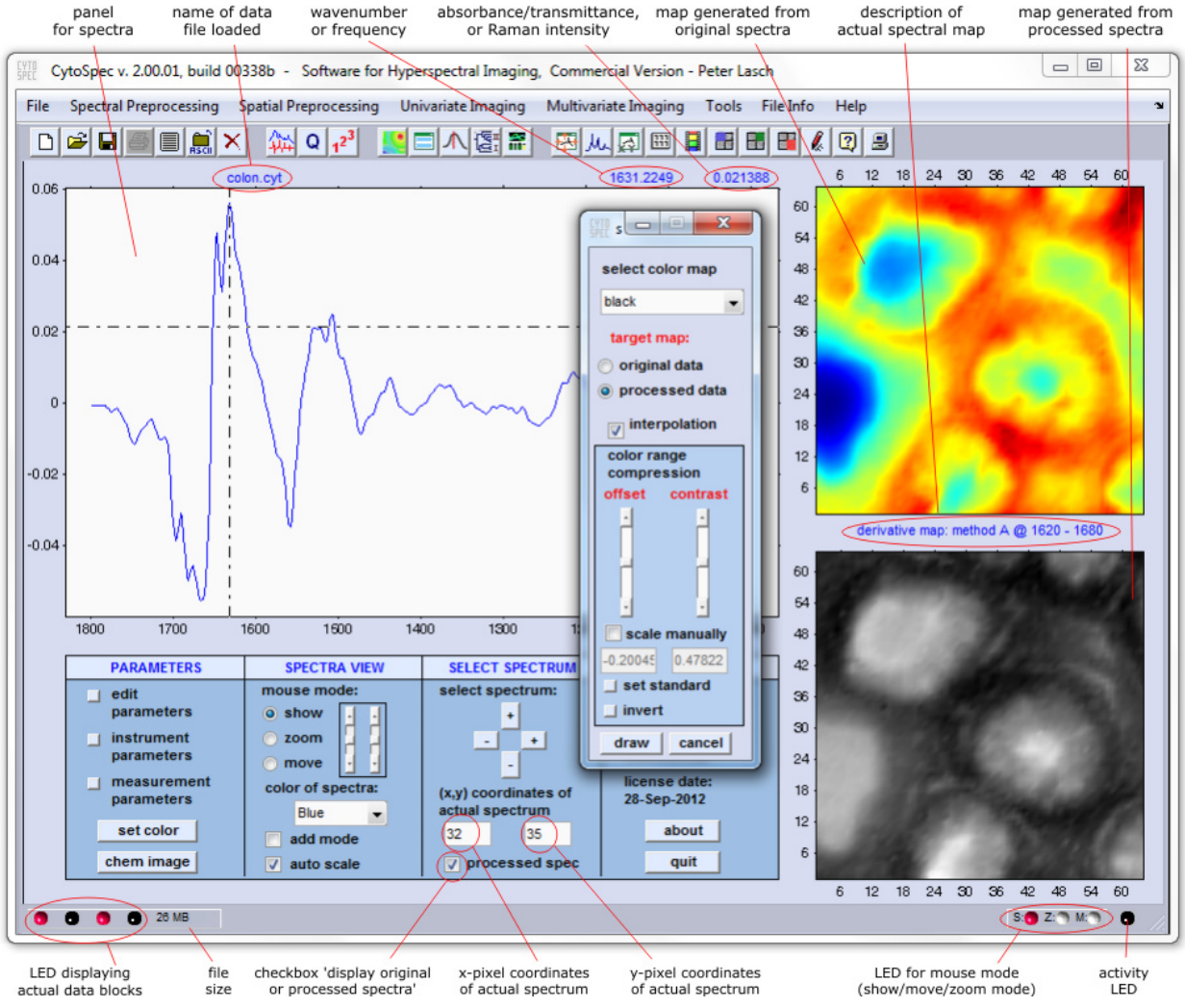
CytoSpec 64-bit: With CytoSpec 1.4.02 a 64-bit version of the program is available (on request, commercial version only). Please refer to the CytoSpec 64-bit documentation for details.

The main window of the program shows three panels where the spectra and hyperspectral maps are displayed (cf. screenshot of the main window of the CytoSpec program below). The two panels to the right are used to display the spectral maps obtained from original spectral data (upper right panel) or from processed spectra such as preprocessed data, derivatives, and (de)convolution data (see lower right panel). The latter panel is used also for displaying hyperspectral images obtained from multivariate spectral analysis methods: (i) clustering and (ii) endmember selection techniques, (iii) principal component analysis or segmentation maps produced on the basis of analyses by (iv) artificial neural networks (ANN).

The color scheme utilized to display spectral maps can be changed (see button '*set color*' of the main window, cf. also the Set colors option from the *tools* pull down menu). In addition, the CytoSpec program offers a lot of useful tools to adjust parameters such as image contrast and color scaling.

The CytoSpec program gives you easy access to the spectra. When clicking into a hyperspectral map (left mouse button) the

- spectrum with the pixel coordinates of the mouse pointer is displayed in the left window,
- the pixel coordinates of the active spectrum appear in editable text boxes, and
- some information of how the spectral map was produced is displayed in the field between both panels to the right.



Many options for displaying spectra can be modified (button '*display options*', option '*display options*' of the *tools* pull down menu). For more details please refer to the chapter ➡ Working with Spectra - Basic Concepts.

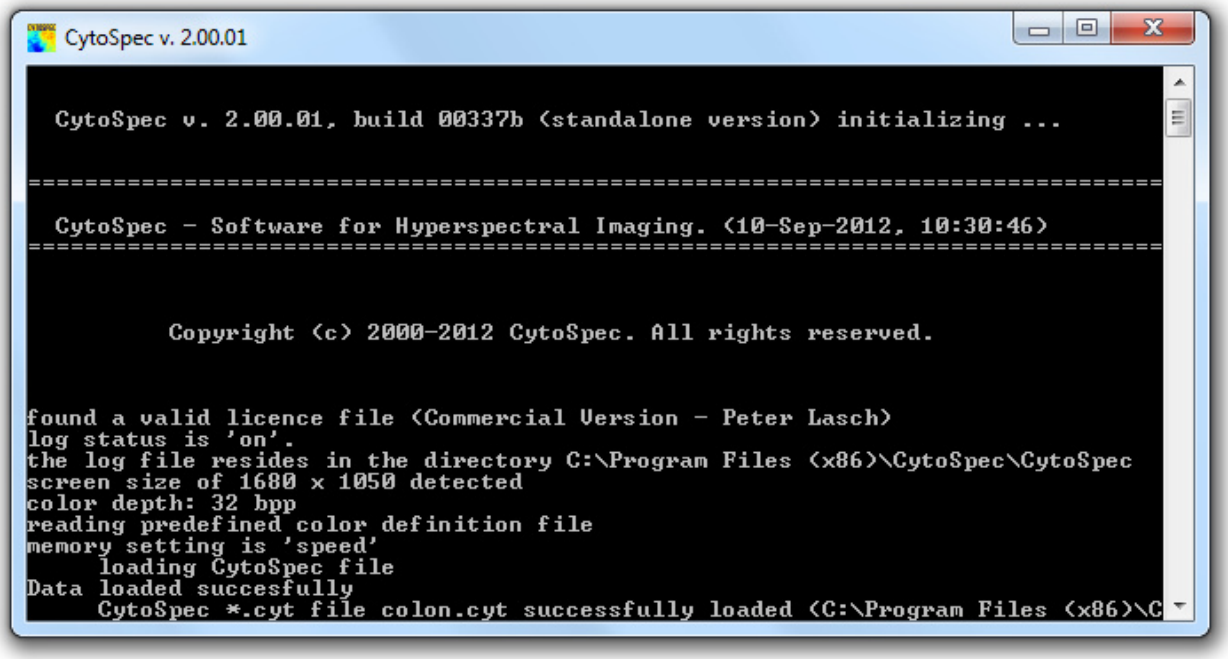
One of the basic principles of this software is to ease data import of spectral data into other applications, and also from other programs into CytoSpec. To give an example, selecting individual spectra for ASCII export can be started by a simple click with the right mouse button (for details see chapter ➡ Export).

When starting the CytoSpec program the main Window and a command line window (see screen shot below) will appear. While the main window allows to interact with the program, the command line window shows input parameters and displays non-standard errors. These messages are stored in a log-file ('*history.log*'), which can be found in following directories:

CytoSpec standalone version (32-bit), Windows operating system (W2k/XP/Vista/W7): root directory of the CytoSpec program, usually C:\program files\CytoSpec\CytoSpec\history.log

CytoSpec pcode toolbox under Matlab (64-bit), Windows operating system (W2k/XP/Vista/W7): Matlab user directory, usually C:\users*CytoSpec-user*\Documents\Matlab\history.log with '*CytoSpec-user*' being the Windows user name

When reporting problems and software failures (bug reports) please send the file '*history.log*' to the following e-mail address: service@cytospec.com



Example of the command line window

This section describes the interactive display mode of the CytoSpec v.2.00.01 program. In this version the following interactive display functions are available:

Mouse modes:

- **show**: displays the pointer's x- and y-position in the '*spectra*' information field
- **zoom**: can be used to interactively enlarge/reduce a part of the spectrum
- **move** or '*roll' mode*': rolls or moves the spectrum in any direction

Spectra of the mapping measurements are displayed in the left panel of the main window. Options for displaying the spectra can be selected from the main CytoSpec window (the window '*display options*' of earlier CytoSpec versions has been removed). The following functions are available:

- display interactively wavenumber and absorbance values (radio button **'show'**)
- interactive expand/compress mode (radio button **'zoom'**)
- interactive shift mode (radio button **'move'**)
- set curves color (list box **'color of spectra'**)
- zoom window horizontally/vertically (sliders)
- display multiple spectra (check box **'add mode'**)
- x-y auto scale for active spectrum (check box **'auto scale'**)

In the mouse mode **'show'** the pointer's x- and y-position will be indicated in the '***spectra***' information field of the main window.

Using the **'zoom'** mouse mode: to zoom in (out) spectra in the display window you have first to activate the appropriate radio button. Then, if you wish to enlarge parts of a spectrum you have to set the first position for the spectrum window by a left mouse click. Hold the button and drag a frame by moving the mouse to the second point. If you release the mouse button, the contents of the frame will be expanded to the full display window. To reduce the size of the spectrum, you can use the right mouse button in the same way.

A left mouse click, that is when no frame is drawn, expands the spectrum in the display window by 200%. A right mouse click will reduce the size to 50%.

The **'move'** mode is used to roll the spectrum in any direction. The display limits are shifted so that all spectra displayed are shifted by the same amount. Again, you have to activate the appropriate radio button before you can start. Spectra can then be moved if you click (left button) into the display window and move the pointer to the desired end point (the mouse pointer will change to a closed hand symbol). Release the left mouse button to finally roll the spectra into the desired direction.

How to select spectra for display? There are two methods available to display spectra of active hyperspectral maps:

Method A: If you click into one of the hyperspectral maps ('*show mode*' activated) to the right the x/y-pixel coordinates within the map are obtained and the respective spectrum is printed. Furthermore, pixel coordinates of the active spectrum are displayed in the text boxes of the main window (cf. x/y coordinates of active spectrum. Depending on the character of the map (map produced from original or processed data) the check box '*show work spectrum*' is activated (or deactivated). If images re-assembled by multivariate imaging (PCA, HCA etc.) were activated CytoSpec displays an error message.

Method B: Alternatively, spectra can be selected by manually editing the edit fields '*coordinates, x or y*' and/or by activating the check box '*processed spec*'. If the corresponding map does not exist (e.g. there is no map of original absorbance spectra), an error message will be displayed.

Hot Keys & Icons

A number of icons is displayed CytoSpec's main window. The following table will give you a short overview about icons and associated CytoSpec functions.

Additionally CytoSpec provides many *Hot Keys* to help speed up image re-assembling and exporting data to other applications. You will find that you can save a lot of time by eliminating the need to move the mouse and to choose options from the pull down menus.

ICONS

 clear all	 load spectral map	 save all
 print	 customize	 export ASCII spectra
 exit	 calculate derivatives	 quality tests
 batch mode	 chemical maps	 chemical movie
 frequency maps	 HCA maps	 Synthon maps
 autoscale spectra	 display spectra	 set display limits manually
 grid on/off	 set color	 capture spectral plot
 capture upper image	 capture lower image	 edit
 help	 about	

HOT KEYS

Control-A	about	**Control-C**	set colors
Control-D	display spectra	**Control-E**	edit parameters
Control-F	frequency maps	**Control-H**	help
Control-I	chemical maps	**Control-L**	load data
Control-M	save Matlab	**Control-O**	customize
Control-P	plot	**Control-Q**	clear all
Control-S	save data	**Control-X**	quit the program
Control-1	capture map from original data (bmp)	**Control-2**	capture map from processed data (bmp
Control-3	capture spectral plot (bmp)	**Control-4**	export mapping data (original data, ASCII)
Control-5	export mapping data (processed data, ASCII)		

Internal data organization: Spectral data are stored in one file consisting of up to four distinct data blocks:

1. Original spectra: the first block is reserved exclusively for the original spectra, usually absorbance or transmission spectra, or Raman intensities in case of Raman maps. This block is usually not modified. Exceptions are the functions '*cut*' and '*interpolate*' of the preprocessing pull down menu and the functions '*swap data blocks*', '*rotate*', and '*flip*' of the '*tools*' pull down menu.
2. Preprocessed spectra: this data block contains the results of data manipulations on original spectra (e.g. preprocessed data).
3. Derivative spectra: the third data block is reserved for derivative spectra.
4. (De)convolution spectra: block number four contains the results of spatial preprocessing: spatial filtering, 3D-Fourier self deconvolution (FSD).

Preprocessing, multivariate analysis, and image re-assembling can be performed by choosing one of the existing data blocks (also called source blocks). The results of data manipulation are stored into so-called 'target data blocks'. Please note that existing target data blocks may be overwritten without warning! Target data blocks may vary for distinct functions (see synopses below).

I. Preprocessing (except functions '*cut*', '*crop*', '*interpolate*', '*binning*', '*derivatives*', '*baseline correction*', and '*ABS <--> TR*'):

source block		target block
original spectra	»»	preprocessed spectra
preprocessed spectra	»»	preprocessed spectra
derivative spectra	»»	derivative spectra
(de)convolution spectra	»»	(de)convolution spectra

II. '*Derivative calculation*':

source block		target block
original spectra	»»	derivative spectra
preprocessed spectra	»»	derivative spectra
derivative spectra	»»	derivative spectra
(de)convolution spectra	»»	derivative spectra

III. '*Cut*', '*crop*', '*interpolate*', '*binning*', '*rotate*' & '*flip*':

source block		target block
original spectra preprocessed spectra derivative spectra (de)convolution spectra		The functions *cut*', '*crop*', '*interpolate*', '*binning*', '*rotate*' & '*flip*' modify the number of points (pixels), or the point (pixel) spacing in the spectral and/or spatial dimensions, respectively. All available data blocks are modified .

IV. Preprocessing: '*baseline correction*' & '*PCA based noise reduction*':

source block		target block
original spectra	»»	preprocessed spectra
preprocessed spectra	»»	preprocessed spectra
derivative spectra	»»	not possible
(de)convolution spectra	»»	not possible

V. '*Swap data blocks*':

source block		target block
original spectra	»»	Data block of original spectra will be overwritten by data block of preprocessed spectra
preprocessed spectra	»»	
derivative spectra	»»	not possible
(de)convolution spectra	»»	not possible

VI. '*ABS <--> TR*':

source block		target block
original spectra		Original spectra will be overwritten by converted data. All other data blocks will be deleted!
preprocessed spectra		
derivative spectra		
(de)convolution spectra		

VII. '*Dispersion correction*':

source block		target block
original spectra		Dispersion correction can be carried out only with transmittance spectra stored in the data block of original spectra. Spectra are dispersion corrected and converted to absorbance spectra. The target data block is the block of preprocessed spectra. Other data blocks are not modified.
preprocessed spectra		
derivative spectra		
(de)convolution spectra		

VIII. '*Subtraction*':

source block		target block
original spectra	»»	preprocessed spectra
preprocessed spectra	»»	preprocessed spectra
derivative spectra	»»	derivative spectra
(de)convolution spectra	»»	not possible

IX. '*Spatial Filtering*':

source block		target block
original spectra	»»	(de)convolution spectra
preprocessed spectra	»»	(de)convolution spectra
derivative spectra	»»	derivative spectra
(de)convolution spectra	»»	(de)convolution spectra

Since CytoSpec version 1.1.04 the presence/absence of data blocks is visualized by four LEDs located in the lower left corner of the Main window. A red LED indicates the presence, a black LED the absence of the respective data block.

Memory management modes: CytoSpec offers two distinct memory management modes. This functionality was introduced to allow highly memory consuming operations also with limited RAM resources. With version 2.00.01 and the availability of the 64-bit version of CytoSpec the memory mode '*compress*' is obsolete and has been removed.

1. **'speed'** the fastest mode, but highly memory consuming. Recommended for HCA with hyperspectral data sets containing up to 128 x 128 pixel spectra (32-bit version, 4 GB of RAM required). In this mode all spectral data are hold in memory with 8 byte precision as float64 values. The option is recommended also for HCA of larger hyperspectral data sets when using the 64-bit version of CytoSpec (Matlab toolbox)
2. **'intermediate'** - relatively fast, but less memory consuming. Try this option when datasets are large and/or the amount of installed RAM is reduced. Spectral data are stored with 4 byte precision as float32 values on disk. Only the required data block will be loaded and is held during the calculations in memory. When performing HCA, the distance matrix will be held in RAM.
3. **'compression'** - this memory option has been removed with version 2.00.01 (see 64-bit version of CytoSpec for details).

With CytoSpec version 2.00.01 (Sep. 2012, built version 338) the window *'display options*' has been removed. The functionality is now an integral part of the main window of CytoSpec.

How to select spectra for display? Three methods available to select spectra for display:

Method A: By clicking into one of the hyperspectral maps ('show mode' is activated) the (x,y) pixel coordinates are obtained and used to display the respective spectrum. The (x,y) coordinates are displayed in the edit fields of the main window (see also (x,y) coordinates of active spectrum: Main Window - Basic Concepts). Depending on the character of the map (map produced from original or pre-processed spectra) the check box ' processed spec' will be checked or unchecked. Note also that CytoSpec cannot display spectra from segmentation maps produced on the basis of multivariate image segmentation methods such as HCA, KMC, FCM, etc..

Method B: Alternatively, spectra can be selected by manually editing the edit fields '(x,y) coordinates of active spectrum' (see Main Window - Basic Concepts) and/or by activating the check box 'show work spectrum'. If the corresponding map does not exist (e.g. there is no map of original absorbance spectra), an error message will be displayed.

Method C: Spectra of the pixel coordinates x±1; (y±1) can be displayed by pressing the buttons '+' or '-' (main windows, option 'select spectrum')

How to Produce Hyperspectral Maps

The basic idea of hyperspectral imaging (or mapping) is to derive for each spectrum of a given set of spatially resolved spectra one single spectral parameter. This parameter is then color scaled and plotted as a function of the spatial (x,y) coordinates. There are a number of techniques, which can be used for hyperspectral mapping. The most frequently used technique is the *so called* chemical mapping technique. In chemical mapping (also referred to as functional group mapping) the absorbance, transmittance, Raman intensity, half-width, or the frequency of a given vibrational band is color encoded and plotted against the spatial (x,y) coordinates. This univariate imaging techniques permits to visualize the spatial distribution of functional groups, or specific chemical substances and is extensively applied in biomedical science, remote sensing and other research areas.

In the CytoSpec program you can create chemical maps by a number of different methods using the absorbance or transmittance, integral absorptions, Raman intensities, ratios etc. These methods are outlined in the chapter Chemical Maps

Furthermore, the CytoSpec software offers the opportunity to re-assemble so-called Frequency Maps. In this method, the maxima of specific bands are obtained and images are re-assembled by plotting the frequency values as a function of the spatial coordinates.

Other imaging techniques are based on multivariate methods of spectral data analysis. The CytoSpec program (v.2.00.01 , September 2012) offers nine different methods of multivariate spectral imaging:

- Principal Component Analysis (PCA) imaging
- Imaging based on hierarchical cluster analysis: HCA imaging
- k-means cluster imaging
- Fuzzy C-means cluster imaging
- Imaging based on ANN (Artificial Neural Network) analysis (SNNS imaging)
- ANN Imaging based on ANN models developed by the help of Synthon's NeuroDeveloper software
- Vertex Component Analysis (VCA) imaging
- n-findr imaging
- Imaging with Distances Values

These methods are extensively described in the respective chapters of the CytoSpec online help.

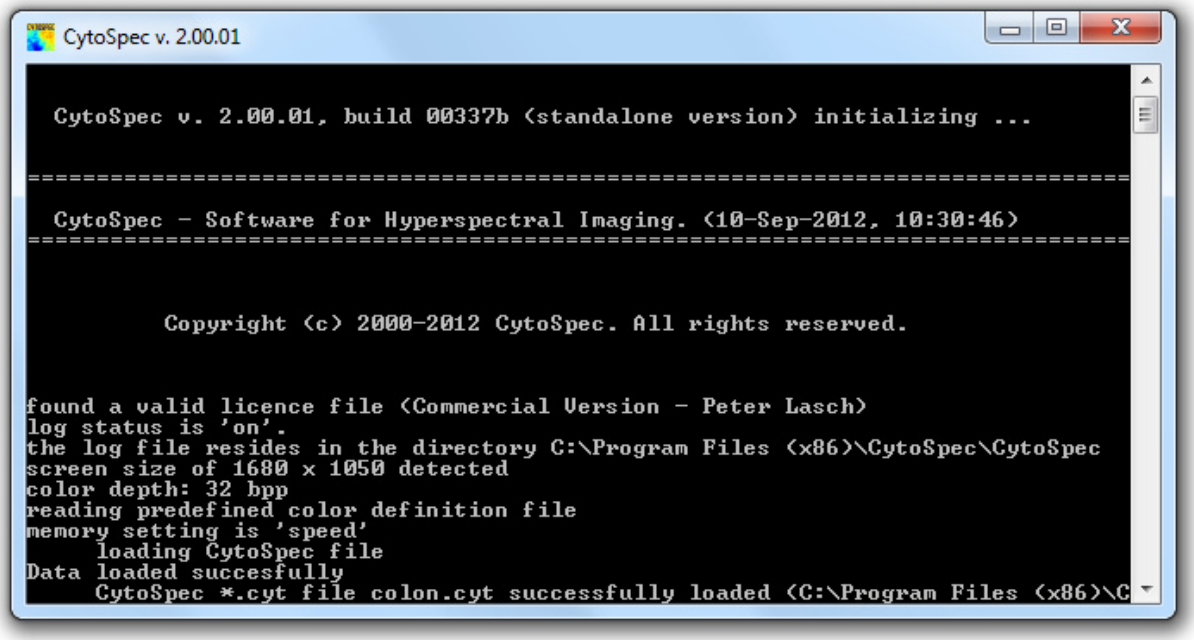
The Log-File

When starting the CytoSpec program the Main Window and a **command line window** (see screen shot below) will appear. While the main window allows to interact with the program, the command line window shows input parameters and displays non-standard errors. These messages are stored in a log-file ('*history.log*'), which can be found in following directories:

CytoSpec **standalone version** (32-bit), Windows operating system (XP/Vista/W7): root directory of the CytoSpec program, usually C:\program files\CytoSpec\CytoSpec\history.log

CytoSpec **pcode toolbox under Matlab** (64-bit), Windows operating system (Vista/W7): Matlab user directory, usually C:\users*CytoSpec-user*\Documents\Matlab\history.log with '*CytoSpec-user*' being the Windows user name

When reporting problems and software failures (bug reports) please send the file '*history.log*' to the following e-mail address: service@cytospec.com



Example of the command line window

How to obtain CytoSpec?

Ordering the full version:

The commercial version (unlimited license, complete functionality) can be ordered from CytoSpec. Please e-mail your request to CytoSpec (e-mail: service@cytospec.com).

Downloading an evaluation (demo) version:

You can also download a free demo version (limited functionality) of CytoSpec. In order to run CytoSpec an additional license key file ('keygen.gen') must be copied into CytoSpec's main directory. This individual license key is NOT included in CytoSpec's zip archive and will be send on request. Please e-mail your request together with your name and an institutional address to the following e-mail address: service@cytospec.com.

Most actual version 2.00.01 - please report software bugs

Download CytoSpec's standalone demo:
version 2.00.01, built 340, March 2013
size: 7156733 byte
CRC checksum: B33D2C94

CytoSpec 64-bit version (Matlab pcode, v. 2.00.01)

Download CytoSpec's 64-bit toolbox:
version 2.00.01, built 340, April 2013
size: 800666 byte
CRC checksum: 6680B172

Test data (binary example files, ASCII data)

Download binary and ASCII test data:

size: 68791131 byte
CRC checksum: 099DC8D0

Earlier demo version of CytoSpec:

Download CytoSpec's demo:
version 1.4.03, built 331b, April 2011
size: 24956334 byte
CRC checksum: 14826283

DOWNLOAD

Please note that you accept with downloading the following license conditions:

You are using the program at your own risk! CytoSpec does not take any responsibility for damages, problems etc. resulting from use of this program. CytoSpec also does not give any warranty for bug-free operation, fitness for a particular purpose or the appropriate behavior of the program.

The software is provided 'AS IS'. For your personal use you can make copies and run as many instances as required, but it is not allowed to further distribute this software.

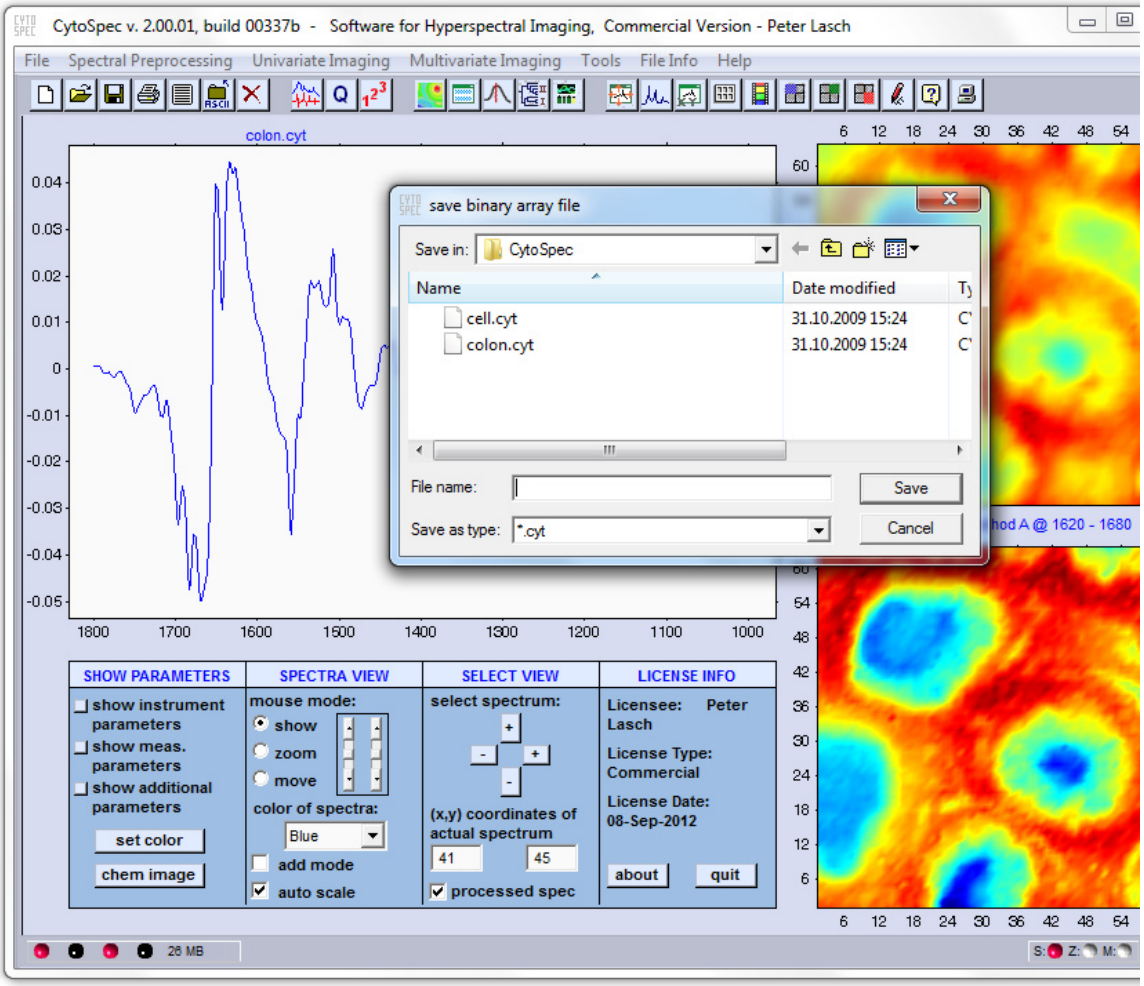
All trademarks mentioned are property of their owners.

LOAD / SAVE DATA FILES

The workspace, i.e. the spectral hypercube containing the original spectral data as well as all processed data (up to 4 different 3D data blocks) and the instrument-, data acquisition-, and additional data parameters can be stored/loaded in form of one single file. This file can be loaded/stored by choosing the 'load' (or 'save') option of the 'file' pull down menu. A standard Windows dialog box allows you to browse the directory structure and to select an appropriate data file name.

To save disk space, data are stored as single precision values (float32) while all calculations are carried out with floating 64 bit point numbers.

Please note: The default file name extension of binary CytoSpec workspace files is '*.cyt'. Files ending with '.dis', '.cls', or '.pca' contain results of HCA, or PCA and have a different format than standard CytoSpec workspace files. These files cannot be loaded by the 'load' function of the 'file' pull down menu. Please note also that CytoSpec's (ver. 1.05.09 and older) older data files ('.mat') or obsolete. To load these files use the 'import' option.



Example: *saving the workspace of a focal plane array detector experiment*.

SAVE MATLAB

This option permits to save the spectral hypercube in a data format that is compatible with Matlab (The Mathworks Inc., Natick, MA).

Save Data:

To store the data file select 'Save Matlab' from the 'File' pulldown menu. A standard Windows file dialog box will appear. Use the dialog box to find a logical location on your disk and enter a filename (e.g. 'fname') with '.mat' as the file extension.

Load data in Matlab:

To open the Matlab file 'fname' cd to the directory where this file resides and type 'load fname' in the Matlab shell. This will cause Matlab to open the file. As a result you will get access to the variable 'C' - the spectral hypercube, a 4D array of double precision or float64 values, and a structure array 'Minfo' containing strings with the file information data.

The array 'C' will have the following dimensions: xdim x ydim x zdim x data block number (1: original data, 2: preprocessed data, 3: derivative data, 4: (de)convolution data, for details see also the chapter Internal Data Organization. The second variable 'Minfo' is a structure array of the following fields:

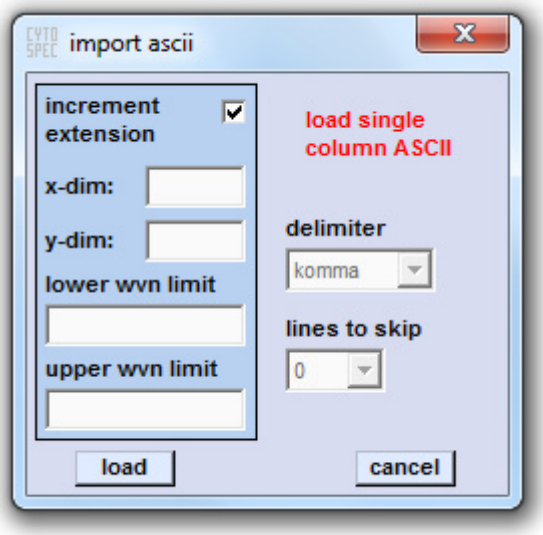
- **Minfo.Readme**: The readme information
- **Minfo.Ver**: Version information of the multispectral data file
- **Minfo.File**: A structure array of three letter code fields such as 'Minfo.File.INS', 'Minfo.File.MAN', Minfo.File.'SRC', etc.). These strings contain important data acquisition parameters and are required for further analyses. The meanings of these parameters are described in the chapter Edit Parameters.
- **Minfo.Org**: A string describing the manipulations performed with the data block of original spectra
- **Minfo.Pre**: A string describing the manipulations performed with the data block of preprocessed spectra
- **Minfo.Der**: A string describing the manipulations performed with the data block of derivative spectra
- **Minfo.Dec**: A string describing the manipulations performed with the data block of (de)convolution spectra

IMPORT SINGLE COLUMN ASCII

Data format: Single column ASCII data contain one column of absorbance/transmittance/Raman intensity data (see example below) from high to low wavenumbers, or frequencies. The user is prompted to enter the high (upper wvn limit) and low (lower wvn limit) value of the IR spectra (in wavenumbers). Upon loading the data, the program determines the wavenumber step increment and the number of data points from the data file.

Single column ASCII data - data structure:

0.0061677871	>> >>	absorbance value at the highest wavenumber
0.0065018409		
0.0068493136		
0.0071548000		
0.0073572793		
0.0074029630		
0.0072586290		
. . . . .		
. . . . .		
. . . . .		
0.0045940578		
0.0041149510		
0.0037795175	>> >>	absorbance value at the lowest wavenumber



xdim & ydim: the dimensions of the map to be imported (in pixel). In the example files (see directory CytoSpecRootDir\Testdata\ASCII\y\) these are xdim x xydim = 11 x 16 (176 spectra). In order to import all ASCII spectra, the files must reside in one single directory. Please make sure that all data files contain the same number of lines.

Wavenumber values: Please indicate the highest (upper wvn limit) and the lowest (lower wvn limit) wavenumber values. Please note that CytoSpec assumes a constant (equidistant) wavenumber step.

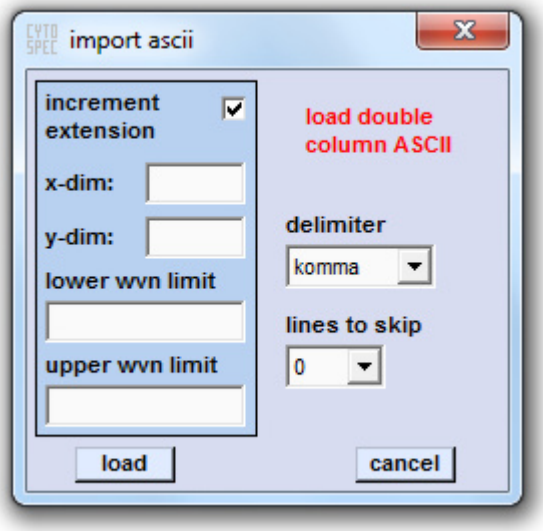
Increment extension: The checkbox marked 'increment extension' determines the complete file names of the files to be read. You can either import single column ASCII data incremented by file extension (e.g. spec.0, spec.1 ... spec.999) or you can import files in which a numeric part of the filename is incremented (for example spec00001.0, spec0002.0, spec0003.0, ... , spec9999.0). For the former option, check the 'increment extension' option box. Up to 1000 spectra can be loaded using this method. If the file number is encoded by the filename, it is possible to load maximally 10.000 spectra. The fields xdim and ydim determine the number of x- and y- data points of the mapping experiment ((x,y) dimensions of the map). They determine which of the consecutively read spectra defines a new row in the map.

After pressing the 'load' button, a standard file window will appear, in which you may browse the directory structure of your computer or network connections. Select the first ASCII data file of the map. A progress indicator shows you then the progress of loading the data files. If problems are encountered during loading of single column ASCII files, try first to import the data of the CytoSpecRootDir/Testdata/ASCII/y/.. directory of the program CD before contacting CytoSpec.

Data format: Double column ASCII data contain the wavenumber/frequency values in the first column and the absorbance/transmittance/Raman intensity values in the second column (see example below, in which data are comma delimited.

Double column ASCII data - data structure:

```
3999.91862793     , -0.00158696
3997.99002686     , -0.00158500
3996.06142578     , -0.00167975
3994.13282471     , -0.00180794
3992.20422363     , -0.00191176
3990.27562256     , -0.00198500
3988.34702148     , -0.00204885
3986.41842041     , -0.00210202
3984.48981934     , -0.00211902
3982.56121826     , -0.00209280
3980.63261719     , -0.00205204
3978.70401611     , -0.00202350
3976.77541504     , -0.00199000
3974.84681396     , -0.00190918
.........         , .........
.........         , .........
.........         , .........
918.01411133      , 0.00025236
916.08551025      , 0.00055385
914.15690918      , 0.00066521
912.22830811      , 0.00064084
910.29970703      , 0.00069492
908.37110596      , 0.00094159
906.44250488      , 0.00125682
904.51390381      , 0.00141472
902.58530273      , 0.00132679
900.65670166      , 0.00110762
```



xdim & ydim: the dimensions of the map to be imported (in pixel). In the example files (see directory CytoSpecRoot\testdata\ASCII\xy\) these are xdim x xydim = 11 x 16 (176 spectra). In order to import all ASCII spectra, the files must reside in one single directory. Please make sure that all data files contain the same number of lines.

Wavenumber values: These options are not available in the double column ASCII import routine as these data are directly obtained from the spectral data files. Please note that CytoSpec assumes a constant (equidistant) wavenumber step.

Delimiter format: You can choose between the following delimiters: comma, space, tab, and semicolon.

Lines to skip: Some data formats may contain header lines (text), which cannot be loaded by the CytoSpec program. To avoid errors upon loading please indicate the number of lines, which should be skipped by the ASCII import routine.

The double column ASCII import function expects files to be incremented by the file extension (e.g. spec.0, spec.1 ... spec.999). This permits to load 1000 double column ASCII spectra at the maximum.

After pressing the 'load' button, a standard file window will appear, in which you may browse the directory structure of your computer or network connections. Select the first ASCII data file of the map and press the 'load' button. A progress indicator shows you then the progress of loading the data files. If problems are encountered during loading of double column ASCII files, try first to import the data of the CytoSpecRootDir/Testdata/ASCII/xy/.. directory of the program CD before contacting CytoSpec. Once the ASCII data are loaded you can ➡ Store them in a binary file format

IMPORT XYZ-ASCII-1

This option permits to load ASCII data of a Raman mapping experiment written in one single multifile. The example below shows the content of a xyz-ASCII file (4 x 4 spectra)

- first column: x-coordinates
- second column: y-coordinates
- first line: wavenumber values of the spectra
- second line: spectrum at the coordinates x(1) y(1)
- separator is tab

		1641.69	1640.83	1639.98	1639.12	1638.27	1637.41	1636.56	1635.7	1634.84	...	702.739	701.726
1	4	0.1245	0.1242	0.1250	0.1263	0.1249	0.1262	0.1241	0.1254	0.1251	...	0.1244	0.1244
1	7	0.1233	0.1243	0.1247	0.1251	0.1275	0.1254	0.1249	0.1249	0.1237	...	0.1245	0.1256
1	10	0.1259	0.1220	0.1242	0.1239	0.1252	0.1225	0.1223	0.1242	0.1245	...	0.1279	0.1240
1	13	0.1243	0.1258	0.1254	0.1225	0.1256	0.1259	0.1261	0.1231	0.1252	...	0.1255	0.1258
3	4	0.1237	0.1218	0.1254	0.1235	0.1244	0.1256	0.1242	0.1247	0.1249	...	0.1240	0.1239
3	7	0.1243	0.1220	0.1254	0.1252	0.1247	0.1246	0.1249	0.1237	0.1234	...	0.1240	0.1264
3	10	0.1248	0.1255	0.1247	0.1211	0.1242	0.1247	0.1246	0.1241	0.1245	...	0.1232	0.1263
3	13	0.1225	0.1235	0.1239	0.1241	0.1237	0.1242	0.1250	0.1241	0.1241	...	0.1237	0.1248
5	4	0.1256	0.1239	0.1252	0.1249	0.1251	0.1254	0.1252	0.1252	0.1241	...	0.1232	0.1244
5	7	0.1236	0.1233	0.1250	0.1231	0.1240	0.1226	0.1234	0.1243	0.1235	...	0.1240	0.1252
5	10	0.1226	0.1246	0.1244	0.1258	0.1225	0.1233	0.1248	0.1237	0.1243	...	0.1255	0.1232
5	13	0.1251	0.1236	0.1247	0.1252	0.1247	0.1266	0.1243	0.1249	0.1248	...	0.1247	0.1266
7	4	0.1266	0.1238	0.1244	0.1252	0.1272	0.1238	0.1260	0.1252	0.1259	...	0.1229	0.1266
7	7	0.1278	0.1248	0.1269	0.1261	0.1246	0.1255	0.1238	0.1249	0.1250	...	0.1254	0.1257
7	10	0.1259	0.1237	0.1238	0.1239	0.1245	0.1243	0.1259	0.1246	0.1220	...	0.1252	0.1237
7	13	0.1233	0.1223	0.1247	0.1232	0.1245	0.1252	0.1241	0.1258	0.1250	...	0.1269	0.1247

After pressing the 'load' button, a standard file window will appear, in which you may browse the directory structure of your computer or network connections. Select the ASCII data file and press the 'load' button. If problems are encountered during loading, try first to import the example files in the directory CytoSpecRoot/Testdata/ASCII/xyz/ (xyz-1.txt - xyz-1d.txt).. before contacting CytoSpec. Once the ASCII data are loaded you can ➡ Save them in a binary file format.

IMPORT XYZ-ASCII-2

The second option for importing XYZ-2 ASCII data of a mapping experiments. The example below shows the content of a xyz-ASCII file. This file contains four columns:

- first column: x-coordinates
- second column: y-coordinates
- third column: ordinate values (wavenumber or frequency positions)
- second line: abscissa values (absorbance, or intensities)
- separator is tab

6.97	-89.09	2000.61	12461.0
6.97	-89.09	1999.65	12585.6
6.97	-89.09	1998.69	12474.8
6.97	-89.09	1997.72	12543.1
6.97	-89.09	1996.76	12616.5
6.97	-89.09	1995.80	12464.2
6.97	-89.09	1994.83	12576.6
6.97	-89.09	1993.87	12432.1
...	...	...	...
...	...	...	...
24.97	-71.09	104.971	84.3432
24.97	-71.09	103.726	74.2909
24.97	-71.09	102.481	54.2039
24.97	-71.09	101.235	18.0652
24.97	-71.09	99.9897	26.0901

After pressing the 'load' button, a standard file window will appear, in which you may browse the directory structure of your computer or network connections. Select the ASCII data file and press the 'load' button. If problems are encountered during loading, try first to import the example files in the directory CytoSpecRoot/Testdata/ASCII/xyz/ (xyz-1.txt - xyz-1d.txt).. before contacting CytoSpec. Once the ASCII data are loaded you can Save them in a binary file format.

IMPORT

CytoSpec offers to import hyperspectral data in various formats, among them in a number of proprietary data formats.

Agilent® (Digilab®)	Hyperspectral imaging data (*.dat) files acquired with the Resolutions Pro software from Agilent®, former Varian® former Digilab® or BioRad®) can be loaded in their native file format. Please note that the *.bsp file (same file name!) must reside in the same directory.
AFMIR	AFMIR data files. Import filter for AFMIR data.
Bruker®	With CytoSpec version 1.03.03 [Feb. 2008] the import filter for Bruker OPUS multifiles has been updated. The filter allows importing OPUS-3D data files, including mapping and FPA data. **Known limitations and restrictions:** Problems may occur when loading data collected with earlier OPUS (OS/2 and pre 3.0 versions). In order to import the data you may have to convert the spectra into an ASCII data format. Problems may also occur if the option '*calculate integral*' (measurements --> special measurements --> mapping) has been selected in the OPUS data acquisition software.
CytoSpec	The import filter for hyperspectral data in the CytoSpec data format .
Horiba Jobin-Yvon®	Horiba Jobin-Yvon data files. The import filter for Horiba Jobin-Yvon ASCII (xyz) data files. Horiba 3D data files can be also imported by CytoSpec in the Galactic spc data data format (Thermo Scientific®). This requires conversion of Raman maps into the spc data format by the manufacturer-specific data acquisition software.
Kaiser®	Kaiser Raman Holospec data files. Import filter for Kaiser Raman Holospec data files.
Matlab®	The import filter for hyperspectral mapping files in a Matlab data format. Details of this file format can be found in CytoSpec's online documentation for the Matlab file format.
NT-MDT®	NT-MDT data files. The import filter for NT-MDT multispectra (Matlab script files).
PerkinElmer®	The filter allows to load data collected with PerkinElmer's SpectrumSpotlight IR data collection software. No problems reported so far.
Renishaw®	An interface to import Renishaw WiRE Raman data files (ASCII only). Renishaw 3D data files can be also imported by CytoSpec in the Galactic spc data data format (Thermo Scientific®). This requires conversion of Raman maps into the spc data format by the manufacturer-specific data acquisition software. When loading the spc data you may have to enter the values for xdim and ydim (number of pixel spectra in x- and y-direction). This information is sometimes not contained in the converted spc data files.
Thermo® (Grams 3D)	Thermo (former Galactic) Grams 3D multifiles (*.spc) can be imported. No problems have been reported so far.
Thermo® (Nicolet)	Thermo Atlus (former Nicolet®) *.map data files. Updated (CytoSpec v. 2.00.01) import filter for Atlus version 7 mapping data files.

WITec® Import filter for WITec confocal Raman microspectrometer data. Requires conversion into an ASCII data format. Please note that CytoSpec assumes the presence of two ASCII text files (e.g. X_fname.txt and Y_fname.txt) in the same directory.

Note that some import filters are still in an experimental stage and may be based on incomplete information of the internal file structures. Bug free operation of these filters cannot be guaranteed! All import filters have been tested using data of the directory "CytoSpecRootDir"/Testdata/bin on the program CD.

Please e-mail data files and the version number of the CytoSpec program you are using to support@cytospec.com when encountering problems with one of the import filters.

EXPORT

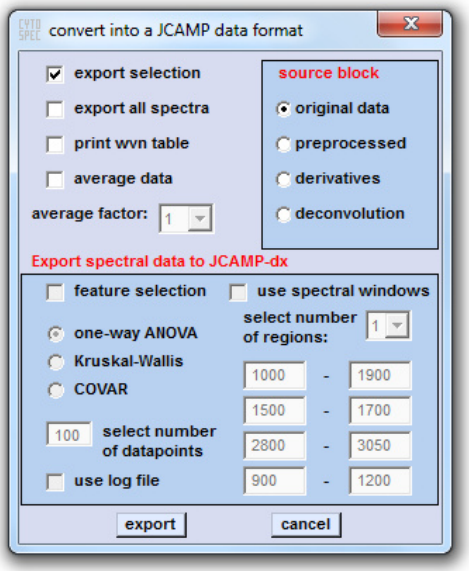
This function permits exporting spectral data to other applications. Spectral data can be exported for re-classification (complete maps) or as a selection of spectra for teaching classification models. The routine offers furthermore an number of options for data reduction such as the selection of spectral windows and an average function working in the spectral domain. The export function offers furthermore specialized routines for automated feature selection such as ANOVA, COVAR, and PCA. Data can be exported in the format for the following applications:

- **SNNS/JavaNNS**: the Stuttgart Neural Network Simulator
- **GA_ORS**: Genetic Algorithm - Optimal Region Selection, a command line based UNIX/LINUX tool for analysis of IR, Raman and MRI data (optimization, feature selection).
- **TOOLDIAG:** TOOLDIAG is a command line based UNIX/LINUX tool for multivariate data analysis (classification, feature extraction, feature selection, etc.)
- **Plain ASCII text files**: simple export function for spectra export to applications software such as OPUS (Bruker Optics), Grams/AI (Thermo), Microsoft Excel, or Origin (OriginLab).
- **JCAMP-dx**: export of spectral data files into the JCAMP-dx data format.

Options of the CytoSpec data export/feature selection routine:

average factor: this factor indicates the degree of reduction of the spectral resolution.

source block: please choose one of the four data blocks: original, pre-processed, derivative, or (de)convolution data.



export selection: if this check box is activated, only a selection of spectra will be exported (details of how to select spectra for export are given below).

export all spectra: a complete 3D spectral data block will be exported. Note that this option is available only in combination with the options 'export plain ASCII' or 'export JCAMP-dx' data files. To export complete hyperspectral maps into a format compatible with SNNS, GA_ORS or TOOLDIAG select the 'use log file' option.

print wvn table: a file 'wvn_conversion_table.dat' is created. This file contains the wavenumber positions of the spectral features.

average data: to lower the computational efforts the data point spacing can be increased by averaging in the spectral domain. Please select an average factor.

feature selection: if activated, you can select between the options ANOVA, COVAR and PCA. Please indicate also the number of data points, that is the number of the best discriminative spectral features. According to the feature selection criterion only the best n features are exported.
Note that only the feature selection option ANOVA has been implemeted yet.

use spectral windows: permits to restrict the spectral information on distinct spectral windows.

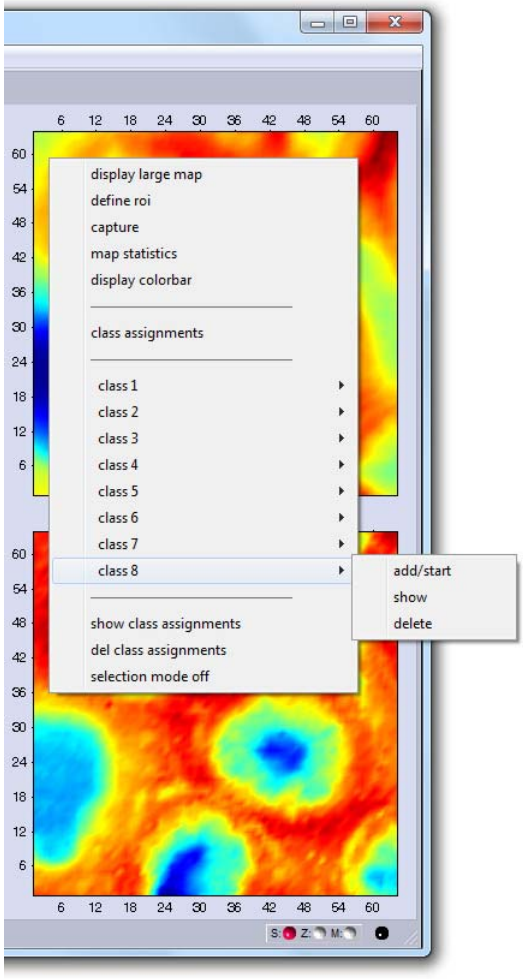
number of spectral windows: please indicate the number of spectral windows

use log file: a log file from a previous CytoSpec export session can be loaded. This function is required for creating external validation (image) tests sets. When this option is chosen the settings for feature selection and spectral windows of a previous CytoSpec export session are utilized. IMPORTANT: after loading the log file the checkbox 'export all spectra' which is routinely inactivated becomes activated. In this way it is possible to export a complete imaging data set for external validation (classification) to the SNNS or ga_ors.

export: pressing this button opens the standard Windows file browser. Please choose an appropriate file name and the target directory.

cancel: the export function is canceled

Either the entire 3D spectral data set, or a selection of spectra can be exported. Spectra can be selected and assigned to defined classes in the following way:



The context menu which is shown in the screen shot ow, can be obtained by a right mouse click over the erspectral maps.

Choose 'class 1' --> and 'add/start' if you want to assign ctra to class 1. Now, you are in the 'select spectra' de. In this mode, the mouse cursor changes its earance (arrow plus cross).

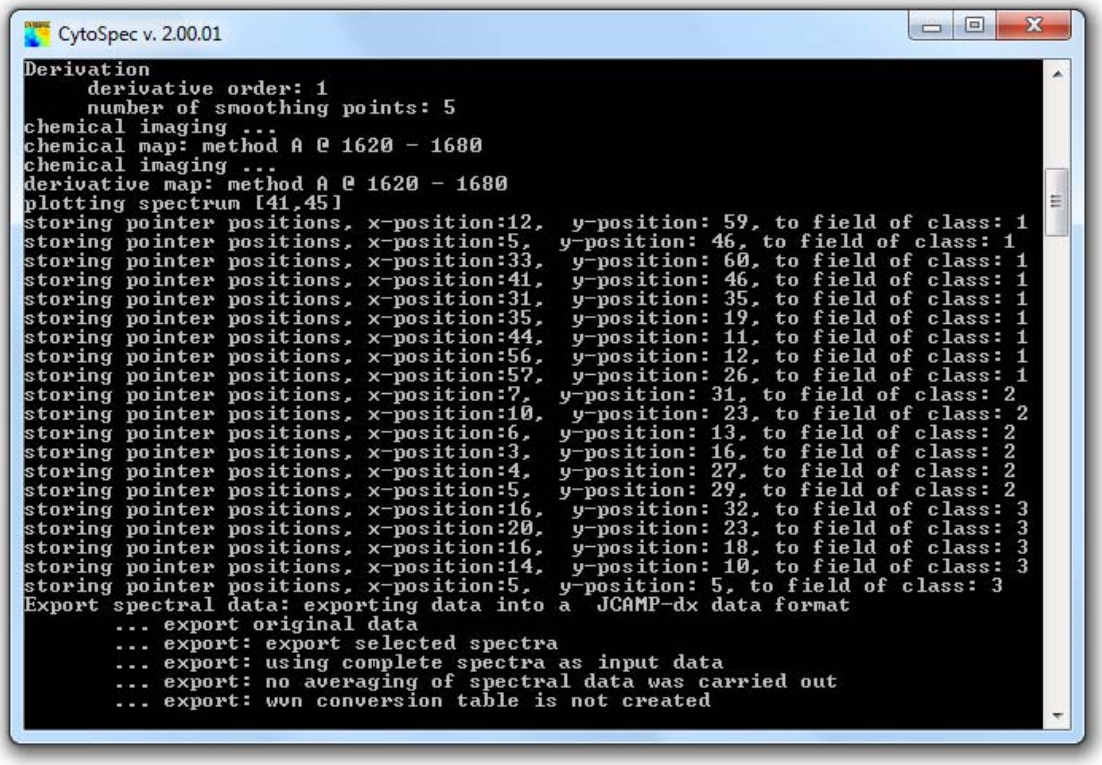
You can select now an unlimited number of spectra by mouse clicks. Please note that spectra are not played in this mode. The spatial coordinates will be en in the command line window.

To assign spectra to another class select 'class X' --> 'add/start'. Then add spectra by mouse clicks.

If all spectra are selected, stop the selection mode by osing 'selection mode off' from the context menu. oSpec returns to the mode 'show spectra' and the use pointer will regain its normal appearance. You can v call the 'export' function and export selected spectra.

leting class assignments: Deleting class assignments be carried out for a class of your choice by the 'class X' and 'delete' function (see figure below); or for all sses by selecting 'delete class assignments' from the text menu.

ow class assignments: To display class assignments an individual class of spectra you can select the 'class X' and 'show' function. Alternatively, you can use the ow class assignments' function from the context menu. The screen output of the 'show class assignments' function is exemplary illustrated in a screen shot of the command line window.



Example of the screen output (command line window) of the 'show class assignments' function available from the context menu of hyperspectral maps.

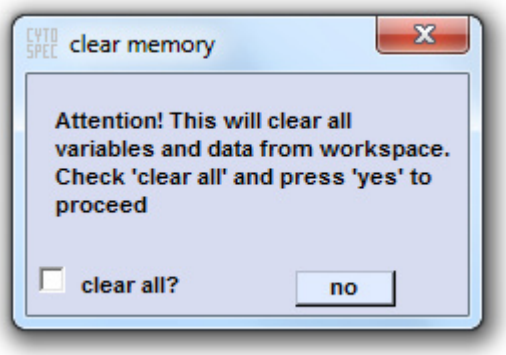
DELETE

This function helps you to organize the spectral data (save disk or memory space). Choose one of the following data blocks that can be deleted:

- preprocessed data.
- derivative data.
- 3D-FSD data.

CLEAR

This function clears the existing workspace and removes all existing hyperspectral maps from the main window. All existing data are lost if they are not stored before. The 'clear'-function may be useful to free memory before performing memory-consuming calculations such as hierarchical clustering (HCA).



PLOT

This function permits to plot the gui into an eps (encapsulated postscript) file. These files are ready to import into vector graphic programs such as CorelDraw.

Alternatively, you can either Capture the spectral maps and store them in an bitmap format or save the zdata information (the data parameters used to generate the surface maps) in an ASCII format by using the Export function of the 'tools' menu. Furthermore, it is possible to produce a screen shot of the program window ('see function Capture')

CUSTOMIZE

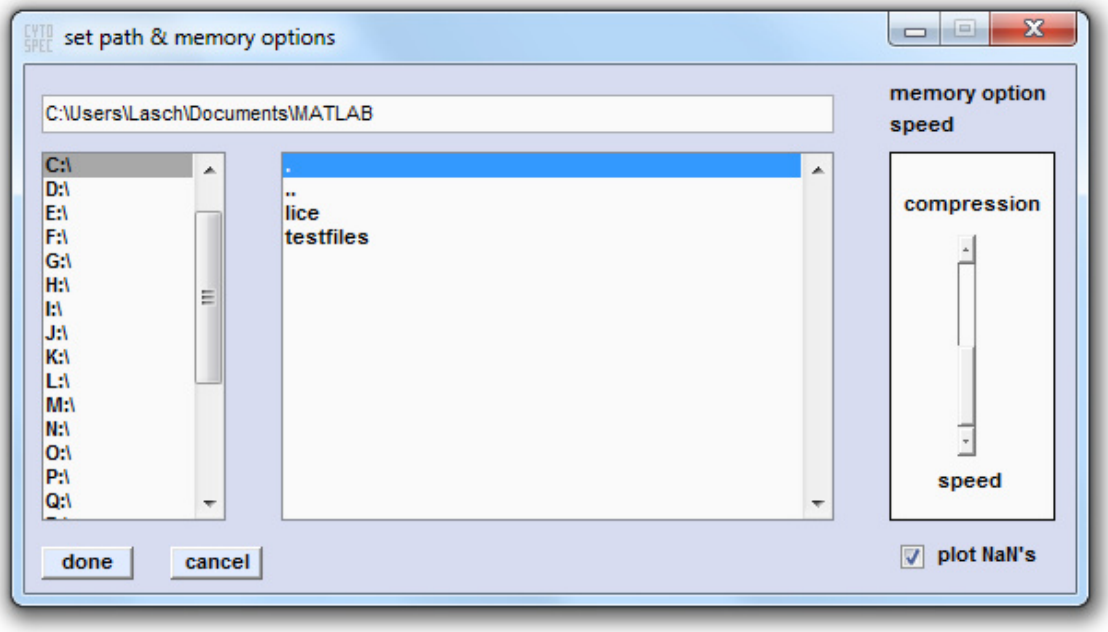
Customize: This function allows to set a permanent path.

Memory management modes: CytoSpec offers two distinct memory management modes. This functionality was introduced to allow highly memory consuming operations also with limited RAM resources. With version 2.00.01 and the availability of the 64-bit version of CytoSpec the memory mode 'compress' is obsolete and has been removed.

'speed' the fastest mode, but highly memory consuming. Recommended for HCA with hyperspectral data sets containing up to 128 x 128 pixel spectra (32-bit version, 4 GB of RAM required). In this mode all spectral data are hold in memory with 8 byte precision as float64 values. The option is recommended also for HCA of larger hyperspectral data sets when using the 64-bit version of CytoSpec (Matlab toolbox)

'intermediate' - relatively fast, but less memory consuming. Try this option when datasets are large and/or the amount of installed RAM is reduced. Spectral data are stored with 4 byte precision as float32 values on disk. Only the required data block will be loaded and is held during the calculations in memory. When performing HCA, the distance matrix will be held in RAM.

'compression' - this memory option has been removed with version 2.00.01 (see 64-bit version of CytoSpec for details).



BATCH MULTIPLE FILES

Batch multiple files. With version 1.2.02 CytoSpec permits to automatically batch-process spectral multifiles. The 'batch multiple files' function is based on the Batch Preprocessing routine and requires a functional predefined macro file (*.cbt -CytoSpec batch)

Before starting the 'batch multiple files' routine it is strongly recommended to carefully check the 'batch preprocessing' function for errors.

Using the function 'batch multiple files':

First, it is required to create a text file which contains path and file name of the batch preprocessing file (first line) followed by paths and file names of the spectral multifiles to be processed (next lines). An example of the file content is given here. This file should have the extension '*.fnm'

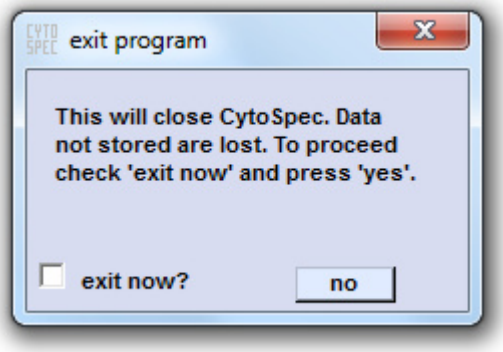
The 'batch multiple file' function can be used to import and process multifiles of many origins. Supported data formats are Matlab (*.mat), Varian (former Digilab) Resolutions Pro, Bruker OPUS, Thermo OMNIC (former Nicolet), Grams/32 (*.spc), PerkinElmer (*.fsm) and CytoSpec (*.cyt).

Spectral data are processed as defined by the batch file order and are stored when finished in the respective file directory by using CytoSpec's own data format and the original file name + '_b.cyt'.

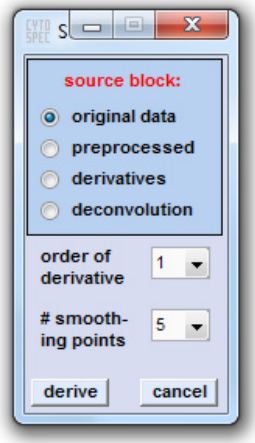
It is not possible to batch-process multivariate imaging functions such as agglomerative hierarchical clustering.

EXIT

Exit: Closes all windows and performs a shutdown of the program. Data not stored are lost



Savitzky-Golay Derivatives



Derivative calculation is carried out by applying the Savitzky-Golay algorithm. In this method n-th order derivatives are obtained while data are smoothed at the same time to reduce the noise. First or second order derivatives can be calculated including 5 to 25 smoothing points. Please note that derivatives are taken in the spectral domain, only. Details of the Savitzky-Golay algorithm can be found in the literature:

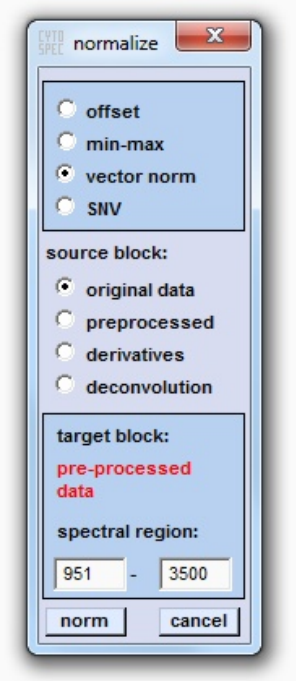
- A. Savitzky and M. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem*. **1964** Vol 36(8):1627.

Any type of data blocks can be handled (including also derivatives). Derivative spectra are stored in a data block reserved exclusively for derivative spectra. If this block is not empty the data are overwritten without warning when obtaining derivatives again (see also Internal Data Organization, Table II). Derivative calculation is always carried out on one complete 3D spectral data block.

Select the source data block by clicking the appropriate radio button, then select the number of smoothing points and the order of the derivative. To finally obtain the derivatives click on the *'derive'* button.

Parameter used for obtaining derivative spectra can be stored within the program workspace and are accessible in the File Info menu *(File Info → File Manipulations → derivatives)*. These parameters are also shown in the command line window.

Normalization



The CytoSpec *'normalization'* subroutine offers four different methods of spectra normalization:

1. Offset correction
2. Min-Max normalization
3. Vector normalization
4. Standard normal variate (SNV)

Offset correction performs are linear correction of the complete spectrum such that at least on point of the spectral region indicated equals zero. Spectra are not scaled in this mode.

Min-Max normalization indicates that all spectra of the source data block are scaled between zero and one, that is the maximal absorbance value of the spectrum in the selected spectral region equals one, the minimum 0. You can use this normalization method to perform a simple band normalization (e.g. for the amide I band).

Vector normalization is carried out in the following way: first, the average value of the absorbances is calculated for the spectral region indicated. This value is then subtracted from the spectrum such that a new average value equals zero. Finally, the spectra are scaled such, that the sum squared deviation over the indicated wavelengths equals one.

$$\sum_k (x_k)^2 = 1$$

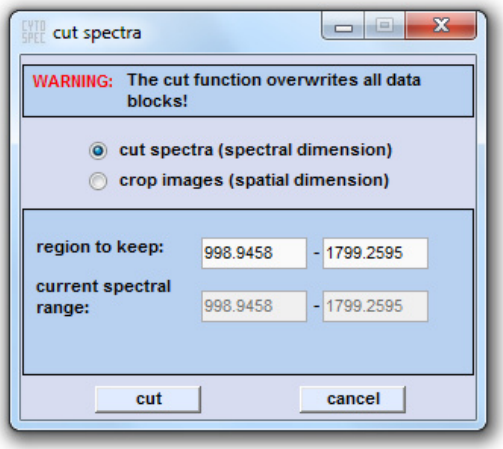
To perform normalization select first the source data block by activating the appropriate radio button (see also Internal Data Organization, Table I). The target data block will be specified in the gray field of the normalization window. Please note that the data of the target data block will be overwritten without warning! Then, select the type of normalization and use the keyboard to enter the wavenumber values between which the spectra the spectra will be normalized. To start normalization click onto the '*norm*' button. If you wish to cancel the operation press '*cancel*'.

Parameters used for normalization such as spectral range are stored within the program workspace and are accessible through the File Info menu (*File Info → File Manipulations → type of data block*). These parameters are also shown in the command line window.

Cut Spectra

CytoSpec's *'cut'* subroutines offer two different methods to cut hyperspectral data sets:

Cutting in the **spectral** domain, and
➡ Crop images in the spatial domains



Cutting in the spectral (z)- dimension can be used to narrow the frequency range of spectral data files. This may be useful to free some memory before memory-consuming calculations such as 3D Fourier self deconvolution are carried out. Define the frequency range to be kept, then click on the 'cut' button to start the function. By pressing the 'cancel' button you can cancel the operation.

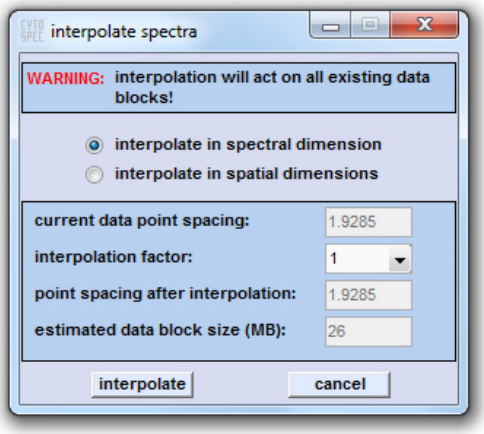
Note that the '*cut/crop*' function overwrites all existing data blocks (see also ➡ Internal Data Organization, Table IV).

The parameter used for '*cut*' are stored within the program workspace and are accessible through the ➡ File Info pull down menu (*File Info* → *File Manipulations* → *type of data block*). These parameters are also shown in the command line window.

Interpolation (Spectral Domain)

CytoSpec's *'interpolation'* routines offers two different methods of interpolation:

interpolation in the **spectral** domain, and
interpolation in the Spatial domains.



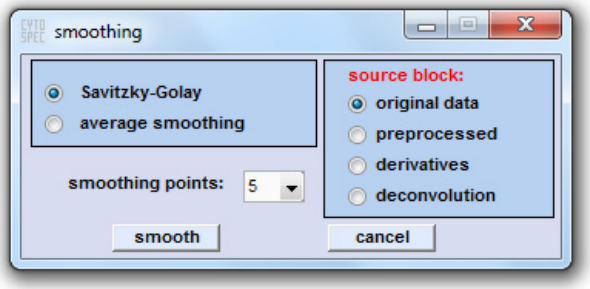
Interpolation in the spectral (z)- dimension changes the spacing between spectral data points. The spacing can be increased or decreased by the 'interpolation factor', which can vary between 1/32 and 32. For example, if a factor of 4 is chosen, the number of data points is increased by a factor of 4, i.e. one frequency interval is filled with (4-1) additional data points. In this case the program performs an one-dimensional interpolation of the spectra. Using a large interpolation factor (e.g. 32) the number of data points of the new spectrum may become rather large. The actual number of data points depends on the start and end frequency and the frequency interval of the original spectrum.

If a factor smaller than 1 is chosen, the data point spacing is decreased. For example, if a factor of 0.25 is chosen, the number of data points is decreased by a factor of 4, i.e. four frequency intervals are merged into one wavelength interval. Consequently, spectral information is lost. Interpolation is useful to reduce the noise or to free some memory before memory-consuming calculations such as 3D-Fourier self deconvolution (3D-FSD) are carried out.

The '*interpolate*' function overwrites all existing data blocks (see also Internal Data Organization, Table III).

The parameter used for interpolate are stored within the program workspace and are accessible through File Info menu (*File Info → File Manipulations → type of data block*). These parameters are also shown in the command line window.

Smoothing



Smoothing: This function smoothes spectra, using either the Savitzky-Golay, or the average smoothing algorithm. Possible values for smoothing points are 5 to 25. Select the source data block as usual, choose the number of smoothing points and click the 'smooth' button to start the operation. Smoothing has a mostly cosmetic effect on the spectrum, reducing the noise at the expense of distorting the signals.

Details of the *Savitzky-Golay* algorithm can be found in the literature:

- A. Savitzky and M. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Anal. Chem. **1964** Vol 36(8):1627.

The parameters used for smoothing are stored within the program workspace and are accessible through File Info menu (*File Info → File Manipulations → type of data block*). These parameters are also shown in the command line window.

Conversion Absorbance ↔ Transmission

TR ↔ ABS conversion: This function performs the conversion from transmission spectra to absorbance spectra and vice versa. Note: The function ABS ↔ TR acts on the complete data block of original spectra and overwrites the data block of original spectra. Furthermore, all other types of data will be deleted.

For the conversion of absorbance spectra to transmission spectra the following formula is used:

$$ABS = \log_{10}\left(\frac{100}{TR}\right)$$

Formula, used to obtain absorbance spectra from transmission spectra:

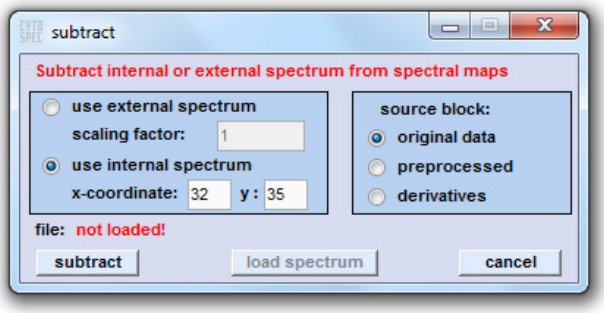
$$TR = 10^{(2-ABS)}$$

Subtraction

Subtraction: This function permits subtraction of spectra from complete spectral data blocks. The function might be useful to compensate for spectral contributions of supporting substrates in transmission type imaging data (example: subtraction of absorbance spectra of thin films).

Spectral subtraction can be carried in two ways: by an internal, or an external spectrum:

- Internal spectrum - a spectrum that is contained in the actual spectral map.
- External spectrum - the spectrum can be loaded (ASCII format).



Using external spectrum for subtraction: in order to use this function check the appropriate radiobutton. Load the external ASCII spectrum (details of data format are given below). Type in the scaling factor and select the source data block. After pressing the '*subtract*' button the external spectrum will be multiplied by the scaling factor and the resulting spectrum is subsequently subtracted from all spectra of the source data block.

Using internal spectrum for subtraction: check the radiobutton '*use internal spectrum*' and choose the (x,y) pixel positions (coordinates) of the spectrum you wish to subtract from the map. Note that upon its initialization the '*subtract*' window will read the actual (x,y) coordinates from the main gui. Press the '*subtract*' button after selecting the source data block.

Source block: Here you can choose the type of the source block for the subtraction function. Please note that 3D-FSD data cannot be used as source data.

Target block: If the source data blocks are of the type original or preprocessed, the target data block will be of the type of preprocessed data. If the source data are of the type derivative the target data block will be also of this type (existing data are overwritten without warning, see also Internal Data Organization). The spectral subtraction routine is always carried out on the complete 3D spectral data block.

Button **load spectrum**: Permits to load a double column ASCII spectrum. If the file could be successfully loaded the directory and the file name is displayed and the button '*subtract*' becomes activated.

Button **subtract**: Starts the subtraction routine immediately.

Button **cancel**: The routine is aborted.

Note: In order to be able to subtract an external spectrum one have first to produce a double column ASCII spectrum (for details of the data format see spectra vap_cut.dat or wap_full.dat; both spectra can be found in the directory CytoSpecRootDir/Testdata/watervap/). Upon loading the external spectrum is automatically adapted such that its data point spacing and its frequency range fits that of the sample data: It will be interpolated (alternative point spacing), cut (broader frequency range), and/or extrapolated (narrower range). Extrapolation is achieved by using the closest absorbance value to fill missing data points.

Dispersion Correction

Dispersion correction: This function performs the correction of dispersion artifacts from transmittance spectra (function written by Dr. Melissa Romeo).

The correction of dispersion artifacts is carried out in the following way:

1. Transmittance spectra are linearly interpolated such that they contain 2^n data points (256, 512, ... , 16384)
2. These spectra are then Fourier-transformed.
3. The first half of the Fourier transformed function is inversely Fourier-transformed. This operation produces a complex function containing a real and an imaginary part.
4. Negative terms in the imaginary component are compensated.
5. Squares of the real and imaginary terms are taken.
6. The squared terms are co-added.
7. The root of the sum yields a phase-corrected spectrum.
8. Transmittance spectra are back-interpolated and converted into absorbance spectra. They are stored in the data block of preprocessed data.

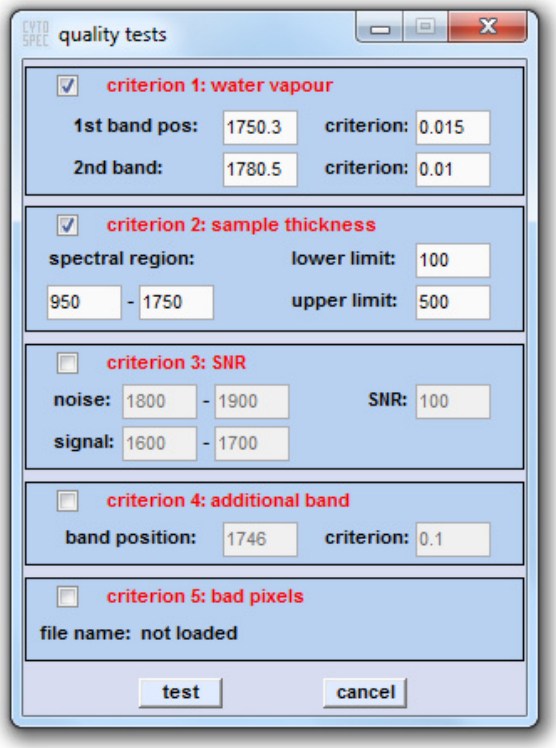
Quality Test: The function 'quality test' implemented in the CytoSpec software comprises five distinct checks for spectral quality:

1. a test for spectral signs of water vapor
2. the check for sample thickness (integrated intensity)
3. the test of the spectral signal-to-noise-ratio
4. a check called '*test for an additional band*'
5. a '*bad pixel*' test (a tool to eliminate spectra from dead pixels of focal plane array detectors

Data organization: the quality tests are performed exclusively on the data block of original spectra. Spectra that have passed the tests are copied without modifications into the data block of preprocessed spectra. Note that existing data of this block are overwritten without warning. If the quality test of a given spectrum is negative, the respective field in the preprocessed data block is replaced by NaN (Not a Number). In this way, spectra tested for poor quality are excluded from further evaluations and will appear in the hyperspectral images as black areas.

If you wish to perform a quality test on preprocessed spectra, for example a sample thickness test after baseline correction, you have to use the Swap Data Block function of the 'tools' pull down menu. This function enables you to overwrite the data block of original spectra by preprocessed spectra.

To enable a test, check the appropriate checkbox and specify the quality test parameters such as absorbance thresholds. Press the '*test*' button to start the quality test or hit 'cancel' if the test should be canceled. The parameters of the test for spectral quality and details of the test results can be found in the File Info menu (File Info → File Manipulations → preprocessed). These parameters are also displayed in the command line window.



1. Test for water vapor:

Sharp water vapor absorption bands can be found in the spectral region between 1300 and 1800 1/cm, a region where many biomaterials exhibit also strong absorption bands. It is therefore recommended to

use water vapor bands above 1750 1/cm for testing. Indicate the precise positions of two water vapor bands which should be utilized for testing and define an absorption threshold criterion. If the absorption of one of the bands is higher than the specified criterion, the test result for the given spectrum will be negative, and the spectrum will be eliminated.

2. Integral absorption as a measure for sample thickness:

The absorbance, integrated over a large spectral region, can be used as a rough measure of sample thickness in transmission type measurements. As many multivariate imaging techniques such as HCA or ANN imaging require a consistent level of the SNR throughout the map, spectra with too low absorptions have to be excluded from further multivariate analysis. On the other hand you may want to eliminate also spectra showing intense signals. This could be the case where the Beer-Lambert law is not obeyed (total absorption, non-linear detector response, etc.)

In order to apply the '*sample thickness*' criterion indicate the spectral region to be used for obtaining the integral. Next, define a upper and a lower threshold for the integral (edit field lower/upper limit). Check the appropriate checkbox to enable the test. A spectrum has failed the sample thickness test if an integration value is determined which is higher or lower than the defined thresholds.

3. Signal/noise ratio (SNR):

This test allows the signal-noise-ratio for individual spectra to be calculated, and to eliminate those that do not fulfill a threshold SNR ratio. Indicate the spectral regions to be used for defining the noise and signal, respectively. For biomedical samples, it is recommended to obtain the signal in the amide I region (1600 - 1700 1/cm) and the noise in the region between 1800-1900 1/cm. Also indicate the SNR threshold and check the checkbox for the SNR test. Spectra are rejected if the SNR is lower than the threshold.

Noise: the standard deviation in the defined spectral range:

$$noise = \sqrt{\frac{1}{(n-1)} \cdot \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

Signal: the maximum ordinate value in the defined wavenumber range

4. Test for an additional band:

This test is useful to exclude spectra from the data set that contain an artifact band (example: regions of a tissue section contaminated by tissue embedding medium). Indicate a typical band position (carbonyl esters of tissue freezing medium: 1746 1/cm) and an absorbance threshold (edit field criterion). Spectra with a higher absorbance at this frequency will be eliminated.

5. Elimination of '*bad*' pixel from FPA data:

Most of the focal plane array (FPA) detectors have so-called '*dead pixels*', i.e. detector elements with zero response to IR radiation. The spectral information at these FPA elements is usually replaced by the camera software with interpolated data from pixel neighbors. If you wish to remove interpolated spectra from the data set, you have to create a simple text file, which should contain the dead pixel (x,y) positions. The text file can be loaded by activating the appropriate check box. Spectra at the given positions are then replaced by NaNs (not a number), i.e. excluded from all subsequent calculations.

Please note: Please use the function Define Spectral Regions to define sample areas in which spectra should be excluded from further analyses.

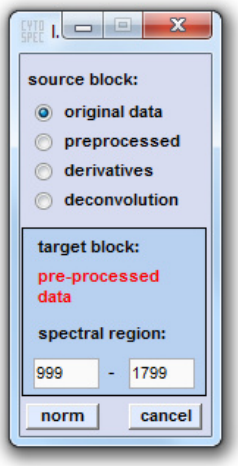
Spectral Baseline Correction

Baseline correction: This function performs a baseline correction. Four different algorithms are available:

- Subtract linear baseline (offset)
- Savitzky-Golay baseline correction
- Baseline correction from curve minima
- Polynomial baseline correction

---

1. Subtract linear baseline (offset): Identical to the offset correction option of the Normalization function. It performs are linear correction of the complete spectrum such that at least one point of the spectral region indicated equals zero. Spectra are not scaled in this mode.



Source block: please select a data block you wish to correct (normalize) by the linear baseline (offset) correction routine.

Spectral region: the spectral region, in which the baseline function is searching for a minimum y-value of the spectrum which is subtracted from the spectrum.

Norm: clicking on the '*norm*' button corrects the baseline.

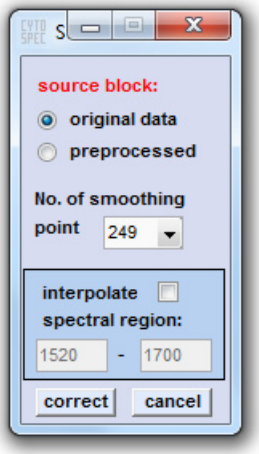
Cancel: closes the application.

---

2. Savitzky-Golay baseline correction: This function can be used to automatically compensate for baseline effects, for instance as a result of scattering. As it is illustrated in the figures below, spectral baseline curves are generated by Savitzky-Golay filtering using a very high number of smoothing points (up to 999).

Baseline corrected spectra are obtained by subtracting the baselines from the original spectra.

Baseline correction can be carried out on original (absorbance/transmittance/Raman intensity) spectra and preprocessed spectra. Please note, that in the latter case existing data are overwritten without warning. Details of CytoSpec's internal data organization can be found in the respective chapter of the CytoSpec online help (Internal Data Organization).



Source block: please select a data block you wish to compensate for non-linear baseline effects.

Number of smoothing points: number of smoothing points used for Savitzky-Golay smoothing.

Interpolate spectral region: a spectral region, in which the slope of the baseline should be interpolated. To activate this feature you have to check the appropriate checkbox and to indicate the wavenumber values of the spectral region you wish to exclude from baseline calculation (in biomedical spectroscopy, this may be the amide I and II region: 1520-1700 1/cm).
Correct: the baseline correction procedure is initiated.

Sancel: closes the application.

Example: The figure below exemplary illustrates how the algorithm of Savitzky-Golay baseline correction works.

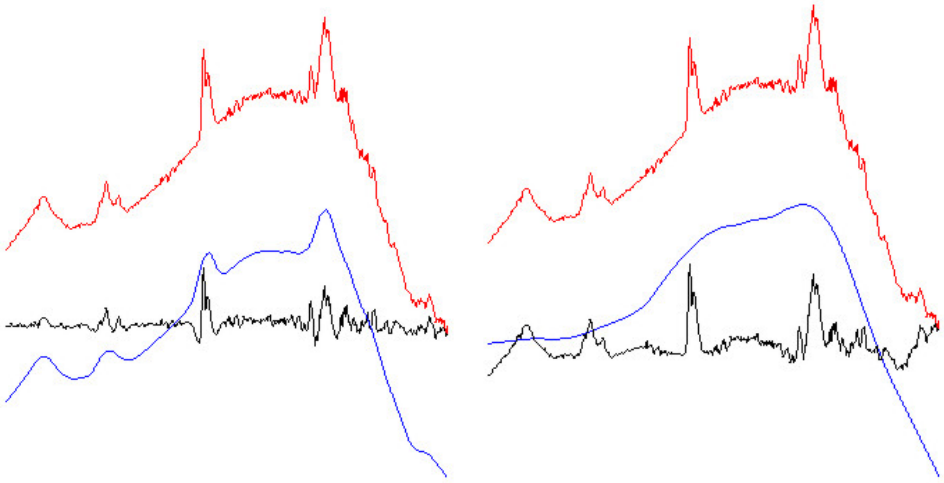
red spectra: original FT-IR absorbance spectra.

blue spectra: baseline curves as obtained by integration of extensively smoothed spectra. The left part of the figure shows baselines obtained with 99 smoothing points and the right panel with 249 points (resolution in the original spectra: 8 1/cm; zero-filling-factor of 4; data point spacing: 2). The example to the right demonstrates additionally the effect of the option '*interpolate region*' which was used to interpolate the baseline in the amide I and II regions (1520 - 1700 1/cm).

black spectra: red (original) minus blue (baseline) spectra. These spectra are stored in the data block of preprocessed spectra.

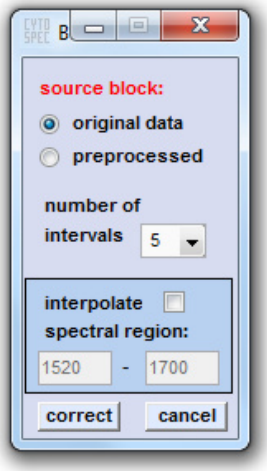
Note: due to the of Savitzky-Golay algorithm, baseline correction might be ineffective in regions close to the upper and lower wavenumber limits (UWN, LWN), particularly if a high number of smoothing points have been chosen. If the number of smoothing points is NOP and the data point spacing is DPS, the baseline correction routine will perform a linear extrapolation of the baseline in the spectral regions.

[UWN] - [UWN-(NOP-1)/2*DPS] and
[LWN] + [LWN-(NOP-1)/2*DPS].



---

3. Baseline correction from curve minima: The function divides the spectrum in segments, or intervals in which minimum y-values (absorbance, Raman intensities) are obtained. These y-values are in the following used to generate a baseline correction curve (by shape-preserving piecewise cubic interpolation) which is subtracted from the original spectrum.



Source block: please select a data block you wish to compensate for non-linear baseline effects. Note that this function does not work on derivative or 3D FSD data.

Number of intervals: number of intervals in which the spectrum is divided.

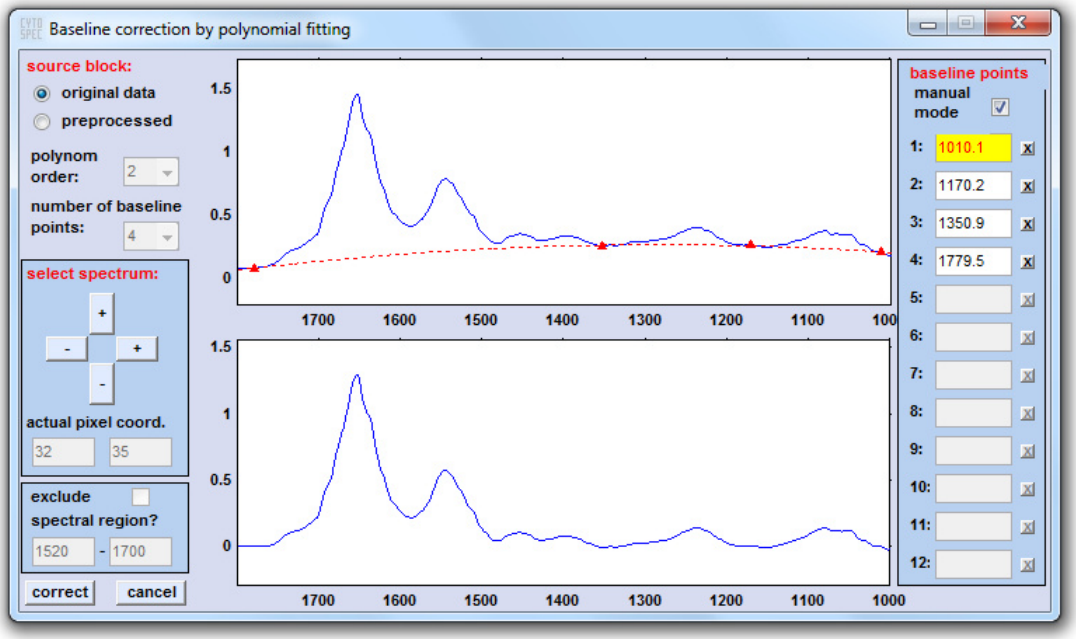
Interpolate spectral region: a spectral region, in which the algorithm should not search for baseline points. If this option is activated you are able to enter the wavenumber values of this spectral region.

Correct: the baseline correction procedure is initiated.

Cancel: closes the application.

---

4. Polynomial baseline correction: This function can be used to subtract a baseline from spectra. The baseline function is a *n-th* order polynom, which is obtained from a set of baseline points that can be defined either automatically, or manually.



Source block: please select a data block you wish to compensate for non-linear baseline effects. Note that this function does not work on derivative or 3D FSD data.

Polynom order: order of the polynom. Valid values are 2-10. Please try to avoid high-order polynoms.

Number of baseline points: select here up 2-12 points which are used to obtain the polynomial baseline function. Note that the number of points should be larger than the order of the polynom.

Select spectrum: the windows to the right display normally the original spectrum with the actual baseline function (upper panel) and the corrected spectrum in the lower panel. The spectrum is read upon initialization of the polynomial baseline function from the main window. If you wish to check the effect of baseline correction on alternative spectra you can increase/decrease the coordinates of the actual test spectrum by pressing one of the four buttons of this panel. The actual pixel spectrum coordinates are displayed in the fields '*actual pixel coord.*'

Interpolate spectral region: a spectral region, in which the algorithm should not search for baseline points. If this option is activated you are able to enter the wavenumber values of this spectral region.

Baseline points, manual mode: allows to manually modify the position of baseline points. Check this checkbox to activate the manual definition mode. If checked one can define baseline points either by mouse-clicks in the upper central panel (shows the original spectrum and the polynomial baseline) or by entering the wavenumber/wavelength values directly in the appropriate edit fields to the right. Note that the field marked by the yellow color will be updated by the next mouse action.

NOTE: each time when the popup menus '*polynom order*' and '*number of baseline points*' are modified the baseline correction function updates all baseline points by a pre-defined algorithm. Baseline points defined earlier may be lost.

x-buttons: when one of these buttons is pressed the respective baseline point is deleted (only possible in the manual mode of baseline point definition).

Correct: starts the polynomial baseline correction procedure.

Cancel: closes the application

Water Vapor Correction

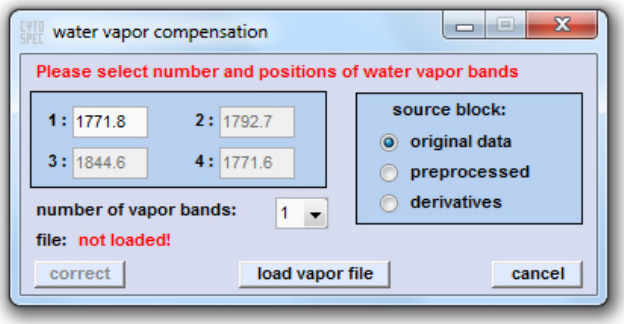
Water vapor compensation: This function permits to automatically subtract a water vapor spectrum from the measurement data such that the spectral effects of water vapor are minimized.

The water vapor correction routine works as follows:

A second derivative spectrum of a pure water vapor absorbance spectrum is obtained.

1. Then, a second derivative spectrum is calculated from the sample spectrum.
2. Depending on your selection, up to 4 separate y-values at defined spectral positions are obtained for both derivative spectra.
3. The water vapor correction factor is calculated by dividing the respective y-values of the water vapor and the sample spectrum. If more than one y-value was selected, the final water correction factor is the average of the ratios.
4. Finally, the sample data are corrected by subtracting the original water vapor spectrum, which was weighted by the water vapor correction factor.

Water vapor correction of derivative spectra: If you wish to perform water vapor compensation on derivative spectra, you have to make sure that spectra are 2nd derivative spectra and that derivative calculations are carried out by choosing 5 smoothing points in the Savitzky-Golay algorithm. The algorithm described above will not work if these two preconditions are not fulfilled.



Number of vapor bands: Please choose the number of water vapor bands on which the spectral compensation for water vapor bands should be carried out.

Edit fields 1-4: Enter the correct positions (in wavenumbers) of water vapor bands. Please note that the band positions may slightly differ from instrument to instrument (calibration) and also as a function of the temperature.

Source block: Here you can choose the type of the source block for water vapor compensation.

Load vapor file: Permits to load a double column ASCII water vapor spectrum. If the file could be successfully loaded the directory and the file name are displayed and the button '*correct*' becomes activated.

Correct: Starts the spectral water vapor correction routine.

Cancel: The routine is aborted.

Data organization (source and target data blocks): Any type of data blocks (except 3D-FSD data) can be handled (including also derivatives). If the source block is of type of original spectra, or preprocessed spectra, the data are stored in the data block of preprocessed spectra. If this block is not empty the data are overwritten. Water vapor compensated derivative spectra are stored in the data block of derivative spectra (existing data are also overwritten without warning, see also Internal Data Organization, Table II). The water compensation is always carried out on the complete 3D spectral data block.

Note: In order to spectrally compensate for water vapor one have first to produce a double column ASCII spectrum of water vapor (for details of the data format see spectra vap_cut.dat or wap_full.dat; both spectra can be found in the directory CytoSpecRootDir/Testdata/watervap/.

Upon loading the external spectrum is automatically adapted such that its data point spacing and its frequency range fits that of the sample data:

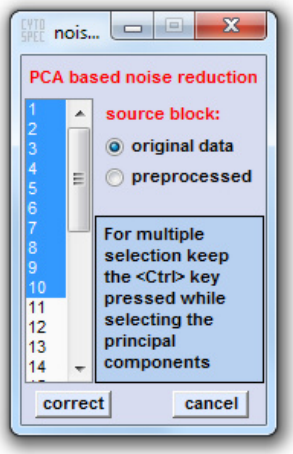
- It will be interpolated (if the point spacing is different), cut (broader frequency range), and/or extrapolated (narrower range).
- Extrapolation is achieved by using the closest absorbance value to fill missing data points.

In the water vapor testdata directory (CytoSpecRootDir/Testdata/watervap/) you can find a test file named 'watervap.mat'. The first data block of this file (original data) contains the original absorbance spectra. Water vapor corrected IR absorbance spectra are found in the second data block of preprocessed spectra. Original spectra are corrected by using the file 'vap_full.dat'.

PCA Based Noise Reduction

PCA based noise reduction: This function can be used to reduce spectral noise. Noise is eliminated by performing principal component analysis (PCA) of the image data and re-assembling of spectra on the basis of a selection of principal components (low order PCs) . In this way, higher-order principal components that are supposed to contain mainly '*noise*' are omitted.

PCA based noise reduction can be carried out on the basis of original or preprocessed data sets. The target data block will be always the data block of preprocessed data.



Important: Please carefully use this preprocessing routine! The decision which of the PCs can be omitted is highly subjective and may cause spectral artifacts.

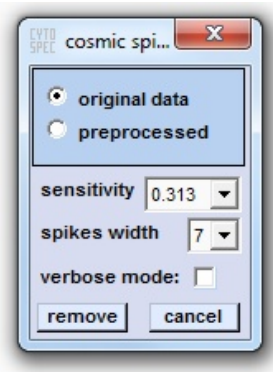
The algorithm has been adapted from a suggestion of Dr. Spragg R. (PerkinElmer) "Addressing Problems in Data Reduction for FT-IR Images of Biological Samples" (Oral Contribution). RISBM - Raman and IR spectroscopy in Biological Medicine. Feb 29-Mar 02, 2004. Friedrich-Schiller-University, Jena, Germany.

Cosmic Spike Correction

Cosmic Spike Removal: The cosmic spike removal tool allows the user to remove cosmic ray features from the spectral (Raman) data. The function can be chosen from the preprocessing pulldown menu. When this function is activated a dialog box shows up which allows the user to change parameters of the cosmic spike removal filter.

With this filter the removal of cosmic spikes is carried out in the following way:

1. Smoothing of the spectral data of choice (original or preprocessed) in the spectral dimension by a 7-point Savitzky-Golay algorithm.
2. A 3D-array of difference values between the un-smoothed and smoothed data is obtained.
3. This array of difference values is normalized by dividing it by the mean standard deviation of the complete array.
4. Cosmic spikes are now obtained by a systematic analysis in the spatial domains. For this purpose, maxima are obtained in each of the image planes of the normalized array of spectral differences. In this context, the parameter 'sensitivity' of the cosmic spike removal dialog box is used to define a threshold above which Raman intensity difference value supposedly indicate the presence of cosmic rays. The higher the sensitivity the lower this threshold.
5. In the next step, the spatial coordinates of spectra with cosmic spike candidates are determined. Cosmic spikes are excised by replacing them with Raman intensities from neighboring frequencies (spectral domain). The parameter 'spikes width' of the dialog box defines the width of the excised spike in points.
6. The cosmic spike removal tool can be applied to data block of original or pre-processed data. In both cases the spike-corrected Raman data are written into the data block of pre-processed spectra. Note that existing pre-processed data are overwritten without warning.



source block: please select the type of data block you wish to correct

sensitivity: used to define a threshold above which Raman intensity difference values supposedly indicate cosmic spikes. The higher the sensitivity the lower the threshold.

spikes width: defines the width of the excised spike in data points.

verbose mode: displays more details of spike removal function.

remove: starts the cosmic spike filter on the data block of choice.

cancel: closes the dialog box.

Batch Preprocessing

Batch preprocessing: This function permits to automize preprocessing of spectral data. When this option is chosen one have to select a predefined macro file (*.cbt -CytoSpec batch) that should be generated before. You can use a plain text editor (Wordpad, etc.) for editing this file. Store your customized *.cbt file in a simple text format (no special characters or format tags!).

CytoSpec's batch processing files contain blocks that start with one of the following three letter codes:

```
DER – derivative
NRM – normalize
CUT – cut
INT – interpolate
SMO – smooth
ATR - ABS --> TR conversion
TRA - TR --> ABS conversion
QAL - quality test
BAS - baseline correction
WVC - water vapor correction
SWA - swap data blocks
CSR - cosmic ray removal
```

Each of the block contains a number of parameters required for processing the imaging data (such as type of datablock, wavenumber values for normalization etc.). These parameters start with a three letter code followed by space character and a numeric value. Details are explained in the example below. Note also that the blocks should be terminated by a 'END' line.

The sequence of preprocessing steps is given by the sequence in the batch file. If you wish to omit preprocessing functions simply comment out the respective block by setting the '#' sign at the first position of a line. Please refer also to the online help or to the example file that comes with CytoSpec's installation CD.

This is an example of the block 'CUT' in a CytoSpec batch (*.cbt) file:

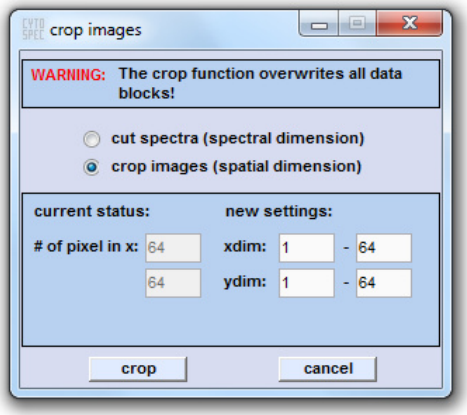
```
# --------------- CUT -------------------------------------------------
CUT
TYP 1    # type of cutting (1-spectral, 2 spatial dimension)
WV1 1000 # first wavenumber for cut in spectral dimension
WV2 1800 # last wavenumber (WV2 larger than WV2!)
XD1 1    # cut, spatial dimension x : first pixel to keep
XD2 10   # cut, spatial dimension x : last pixel to keep
YD1 1    # cut, spatial dimension y : first pixel to keep
YD2 10   # cut, spatial dimension y : last pixel to keep
END
# some lines with comments may follow

next block ...
```

A detailed example of a CytoSpec batch file is given here: ➡ first.cbt

Crop Images

CytoSpec's *'cut'* subroutines offer two different methods to cut hyperspectral data sets:

- Cutting in the spectral domain, and
  Crop images in the spatial domains.



Cropping in the spatial (x,y) dimensions reduces the number of pixel spectra. The number of pixel spectra can be defined by typing the spatial (pixel) coordinates of a rectangular region in the current map.

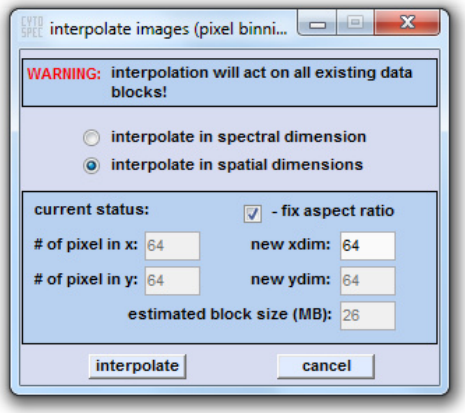
Note that the '*cut/crop*' function overwrites all existing data blocks (see also Internal Data Organization, Table IV).

The parameter used for '*cut*' are stored within the program workspace and are accessible through the File Info pull down menu (*File Info --> File Manipulations --> type of data block*). These parameters are also shown in the command line window.

Interpolate in the Spatial Domain - Pixel Binning

CytoSpec's *'interpolation'* routines offers two different methods of interpolation:

interpolation in the spectral domain, and
interpolation in the spatial domains.



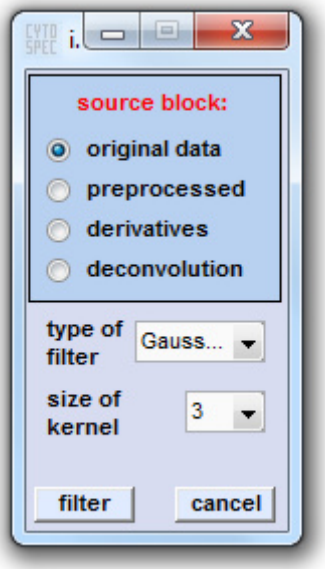
Interpolation in the spatial (x,y) dimensions (pixel binning) changes the number of pixel spectra. The number of pixel spectra can be increased or decreased. For example, if the current number of pixel spectra is 64 x 64 ('*# of pixel in x/y*') and the settings for '*new xdim/ydim*' are 32 x 32 the program performs a two-dimensional interpolation in the spatial domains decreasing the map size by a factor of 2 x 2. Note that map dimensions can be changed to any sizes, e.g. spatial interpolation from an initial map size 64 x 64 to a 30 x 32 pixel map will work. Note also that enlarging map sizes by spatial interpolation may result in very large data files.
Please uncheck the checkbox '*fix aspect ratio*' if you wish to change the pixel ratio between x and y-dimensions.

The '*interpolate / pixel binning*' function overwrites all existing data blocks (see also Internal Data Organization, Table III).

The parameter used for interpolate are stored within the program workspace and are accessible through File Info menu (*File Info --> File Manipulations --> type of data block*). These parameters are also shown in the command line window.

Filter Images

The function '*filter images*' is a function that can be used to apply frequency filters in the xy-image domain. Frequency filters such as low or high pass (smoothing/sharpening) filters are applied in the image domain to smooth, or sharpen chemical images. The target data block of filtering in the image domain will be the data block of (de)convolution spectra if the function is applied to original, or pre-processed spectra. If derivative data are filtered, the target data block will be the block of derivative spectra (see also Internal Data Organization, Table IX).



Select the source data block by clicking the appropriate radio button, then select the spatial filter function and the size of the kernel. To finally apply the spatial filter function click on the *'filter'* button.

Available filters:

- Gaussian
- Mean
- Disk
- Savitzky-Golay smoothing
- Laplacian

Kernel size: indicates the size of the kernel filter function in pixels

The parameter used for '*filtering images*' are stored within the program workspace and are accessible through the File Info pull down menu (*File Info --> File Manipulations --> type of data block*). These parameters are also shown in the command line window.

3D Fourier Self-Deconvolution (3D-FSD)

Basics of 3D-FSD and references to the literature:

This function allows Fourier self-deconvolution in the two spatial dimensions (x and y) and the spectral/frequency dimensions (z) to be performed at the same time. The method can be applied to enhance the spectral and/or spatial resolution and to increase the image contrast. Adapted from the 1-D algorithm described in:

- J.K. Kauppinen, D.J. Moffat, H.H. Mantsch, D.G. Cameron, Fourier Self-Deconvolution: a method for resolving intrinsically overlapped bands. *Appl Spectrosc*. 35(3) **1981**. 271-276.
- J.K. Kauppinen, D.J. Moffat, H.H. Mantsch, D.G. Cameron, Self-deconvolution and first order derivatives using Fourier transforms. *Anal Chem*. 53(9) **1981**. 1454-1457.
- J.K. Kauppinen, D.J. Moffat, H.H. Mantsch, D.G. Cameron, Noise in Fourier self-deconvolution. *Appl. Opt*. 20(3) **1981**. 1866-1879.

Procedure of 3D-FSD:

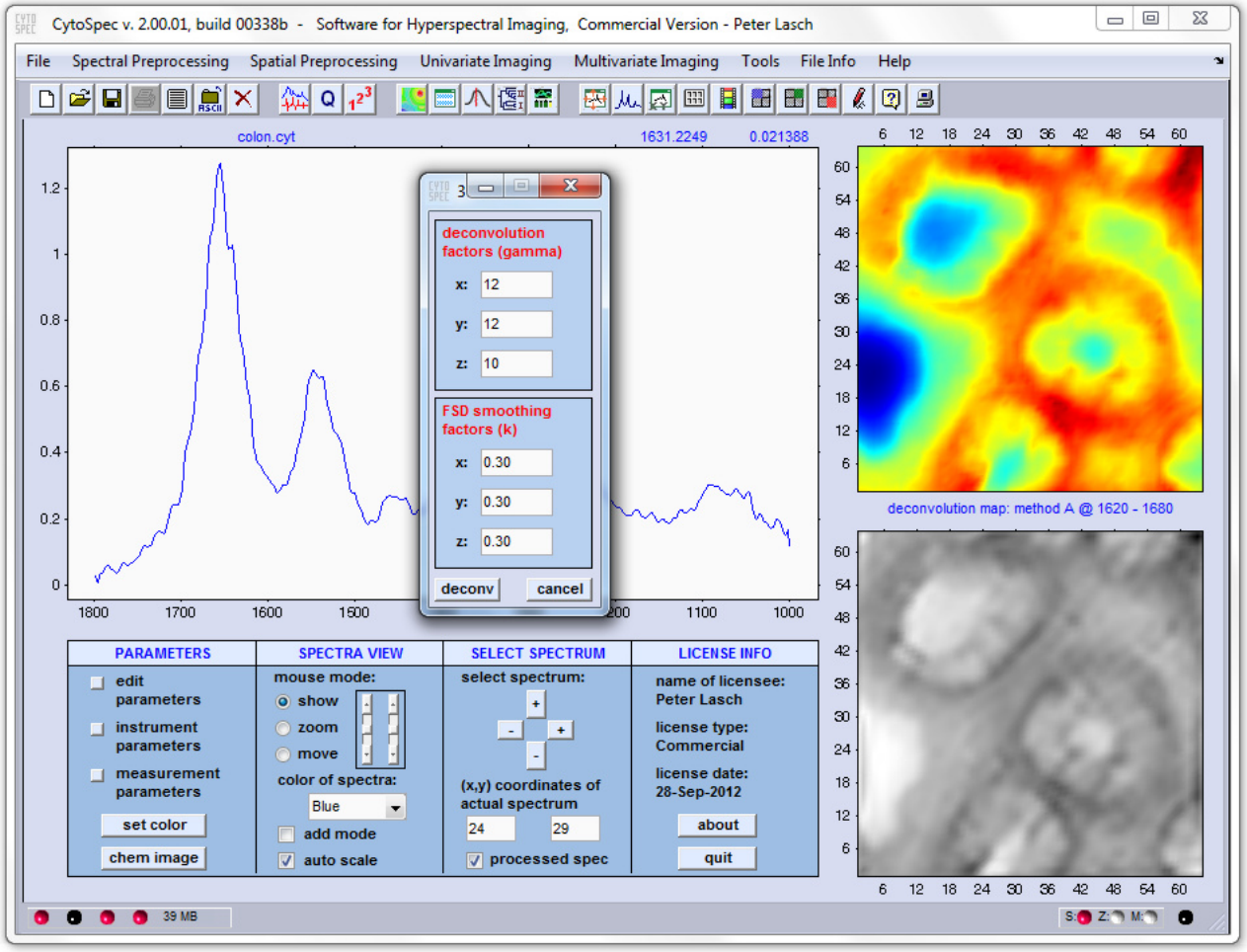
- Spectra are baseline corrected.
- A 3D exponential curve, normalized to the point spacing, is generated.
- A 3D Bessel smoothing function is generated.
- The 3D data file is Fourier transformed.
- The result is multiplied by the product of the 3D exponential and smoothing functions.
- Finally, the result is inverse Fourier transformed to get the deconvoluted data file.

Constraints:

The k-factor must lie between 0 and 0.95: then larger k, then more the data are smoothed. You can apply different smoothing factors for the spatial (k factor x or y) and spectral dimensions (k factor z).

Gamma (the half width at half peak height, gamma > 0) is a parameter of the exponential part of the deconvolution function. The higher gamma, the higher the power of deconvolution.

The peak width value reported by this program actually refers to the REDUCTION in the width of the original peaks in the data. Therefore, a reported width value of 8 means that a peak of approx. width 10 before FSD, will have a width of 2 afterwards.



Example: This example shows the results of 3D-FSD (data acquired by the use of a 64 x 64 mid-infrared MCT focal plane array detector). For chemical imaging, the absorbance values at 1731 wavenumbers of the original spectral (upper plot) and of the FSD data block (lower panel) were color encoded and plotted as a function of (x,y) position. The panel to the left displays a FSD spectrum (blue).

Important:

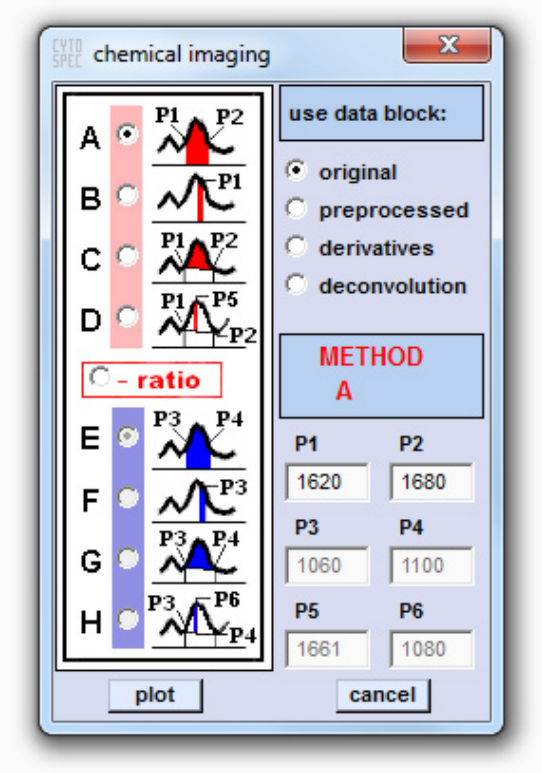
3D Fourier self-deconvolution can be performed only on the original data (data block 1). This function is only suited for envelopes much broader than the spatial/spectral resolution. Avoid oscillatory patterns due to over-deconvolution!

Reference to the literature:

- Lasch P. Naumann D. Spatial Resolution in Infrared Microspectroscopic Imaging of Tissues. Biochim Biophys Acta (BBA) - Biomembranes **2006** 1758(7):814-29.

Chemical Imaging (or functional group mapping): This function permits to produce chemical images, which display defined spectral parameters such as absorbance values at a given frequency as a function of the spatial position. Spectral parameters are color encoded according to the selected Colormap. In the CytoSpec program you can use the following methods for chemical imaging:

- Method **A**: calculates the integrated area in the region P1, P2 (no baseline correction)
- Method **B**: obtains the absorbance value at a given frequency P1
- Method **C**: calculates the integrated area in the region P1, P2 (trapezoidal baseline correction)
- Method **D**: obtains a baseline corrected absorbance value at a given frequency P5. The baseline is obtained from the points P1 and P2
- Methods **A-D / E-H**: calculate ratios using any combination of methods for two regions or absorbances at a given frequency



1. Select first the imaging method. You can choose between methods A-D by activating the appropriate radio button. If you wish to use ratios for chemical imaging you have to additionally select the 'ratio' button and the method in the denominator (methods E-G).
2. Next, type the wavenumber values P1-P6 used as integration borders, or frequency points for construction baselines. Note that the number of values depends on the integration method.
3. Select the data block on which chemical imaging should be carried out. The following data blocks can be used for chemical imaging: original data (i), preprocessed data (ii), derivatives (iii), and 3D-FSD data (iv).
4. Pressing the 'plot' button calculates the respective parameters from the selected spectral data block and plots these parameters as a function of (x,y) position within the map.
5. Button 'cancel': The window for chemical imaging is closed.

Whereas the original data will be displayed in the upper right panel of the main window, chemical images re-assembled from preprocessed, derivatives, and 3D-deconvolution data are given in the lower right panel. Hyperspectral maps produced from multivariate imaging approaches (cluster, PCA, or ANN analysis) are also displayed in the lower right panel of the main window.

A mouse click (left button) into chemical maps results in the following events ('show mode' is active):

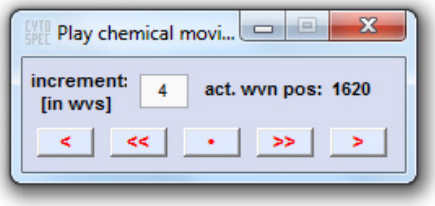
1. an IR or Raman spectrum with the (x,y) coordinates of the mouse pointer is displayed in the left (spectral) panel,
2. the pixel coordinates of the active spectrum appear in editable text boxes, and
3. information of how the spectral map was produced is displayed in the field between both panels (see also chapter Main Window - Basics Concepts.)

related topics

- Manipulating the colormap used for imaging (change colormap, change color contrast/offset)
- How to save spectral images as bitmaps?
- How to export the map data (absorbance/transmittance/Raman intensities) as ASCII-tables?
- How to produce hyperspectral maps (Basics)
- Working with vibrational spectra - basic concepts

Chemical Movie

The function 'chemical movie' can be used to play an animation of chemical images, or frequency slices. With this function the system automatically steps through a range of frequency values, facilitating judgement of the spatial distribution patterns of individual spectral modes. Chemical images are generated by using option '**B**' of the function chemical imaging.



<	one step (frequency slice) backward
<<	click this button to observe an animation of the frequency slices (='*chemical movie*') in backward direction.
°	stop playing the chemical movie
>	one frequency slice forward
>>	plays the chemical movie in forward direction

act wvn pos: the (rounded) frequency position of the actual chemical image in wavenumber units.

increment [in wvs]: This allows to modify the the step size, in wavenumber unit steps (wvs), between frequency slices. If a value of '1' is indicated the step size of the animation will be 1 * '*wvs*', with wvs being the wavenumber step (the point spacing between two frequency slices, see WVS - wavenumber step). The default value for '*increment*' is '4'.

Frequency imaging: This function permits visualization of peak positions, and their variations, within hyperspectral maps. Band positions - either maxima or minima - are obtained and plotted as a function of the spatial coordinates.

Band positions can be obtained from all types of data blocks, including also derivative spectra (for details refer to chapter Internal Data Organization).

Two different methods of the 'frequency map' routine are available: Peak maxima/minima can be obtained from spectra contained in one of the four data blocks, or for overlapping bands from second derivative data blocks (this may also be second derivatives from spectra of the derivative data block!).

The use of second derivatives for peak picking may be useful for the detection of peaks in the presence of strongly overlapping signals, i.e. when the band is only a small shoulder on a strong signal. To some extent 2nd derivatives compensate also for baseline effects. Derivatives should be used with care as noise is considerably amplified.

The **algorithm** of the peak picking routine works as follows:

A. *If the option 'obtain peak positions from derivatives' was NOT selected:*

Spectra of the data block of your choice (see option 'use data block') are interpolated. For interpolation, the spline method is used. Interpolation is carried out in the frequency range indicated in the edit fields 'select spectral region for peak search'. Furthermore, a factor of interpolation can be selected. This factor indicates how many times the number of data points will be increased by interpolation (only in the spectral region selected for peak picking).

If the option 'search maxima' was checked, the algorithm is then searching for the x-positions (frequencies) of the maxima within the spectral region indicated by the user. If minima are chosen, the program is searching for minima. IMPORTANT: If band positions are obtained from the data block of derivate spectra, maxima appear in second derivative spectra as minima and vice versa. There is no check for i) the order of the derivative and consequently ii), no compensation for the inversion of maxima and minima!

The frequency values of the maxima/minima are color scaled and plotted as a function of the spatial coordinates. If you wish to further analyze the band positions by other programs you can access the data matrix of frequency values by using the Export Maps function.

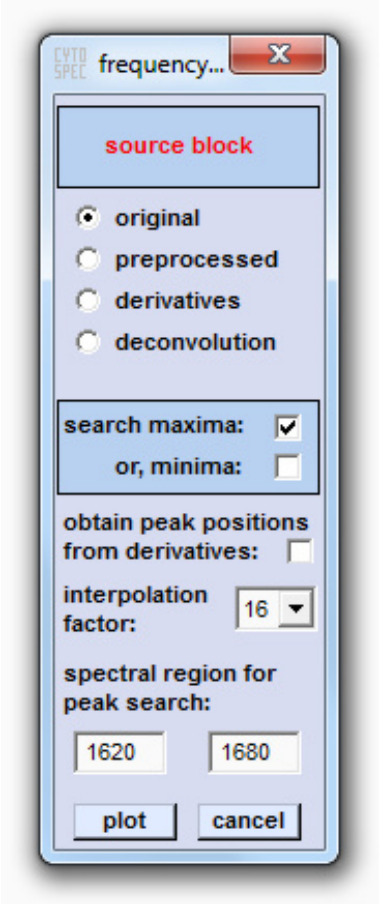
B. *The option 'obtain peak positions from derivatives' was checked:*

Second derivative spectra from the data block of your choice (see option 'use data block') are calculated by applying the Savitzky-Golay algorithm with 5 smoothing points (see also chapter Calculation of Derivative Spectra).

Derivative spectra are interpolated. For interpolation, the spline method is used. Interpolation is carried out in the frequency range indicated in the edit fields 'select spectral region for peak search'. Furthermore, a factor of interpolation can be selected. This factor indicates how many times the number of data points will be increased upon interpolation in the spectral region selected for peak picking.

If the option 'search maxima' was checked, the algorithm is then searching for the x-positions (frequencies) of the maxima within the spectral region indicated by the user. If minima are chosen, the program is searching for minima. IMPORTANT: If band positions are obtained from the data block of derivate spectra, maxima appear in second derivative spectra as minima and vice versa. There is no check for i) the order of the derivative and consequently ii) no compensation for the inversion of maxima and minima!

The frequency values of the maxima/minima are color scaled and plotted as a function of the spatial coordinates. If you wish to further analyze the band positions by other programs you can access the data matrix of frequency values by using the Export Maps function.



search maxima or minima: indicate whether you want to search for peak maxima or minima.

obtain peak positions from derivatives: if this checkbox is checked, second derivatives of the chosen data block will be calculated. In this case, maxima or minima will be obtained from second derivate spectra. Otherwise (if box is NOT checked), peak positions are obtained directly from the data block chosen.

interpolation factor: this factor indicates how many times the number of data points will be increased by the interpolation.

select spectral region for peak search: the spectral region which will be used for peak picking.

use data block: please choose the appropriate data block on which peak picking should be carried out

plot: the peak picking routine is started.

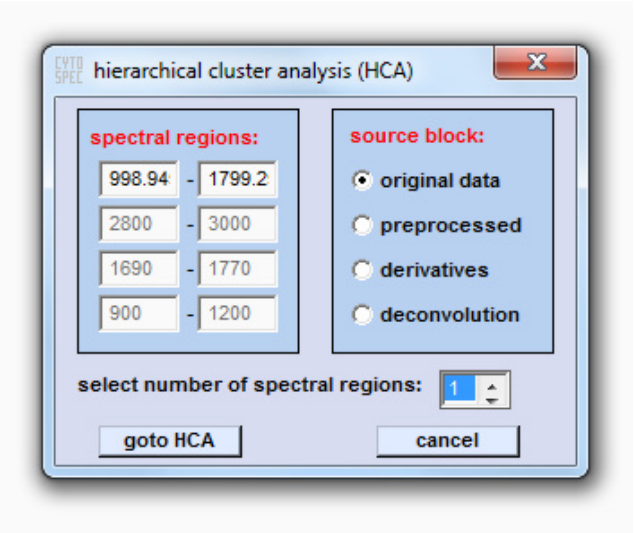
cancel: the '*frequency plot*' window will be closed.

Related topics

- Manipulating the colormap used for imaging (change colormap, change color contrast/offset)
- How to save spectral images as bitmaps?
- How to export the map data (absorbance/transmittance/Raman intensities) as ASCII-tables?
- How to produce hyperspectral maps (Basics)
- Working with vibrational spectra - basic concepts

Create HCA Segmentation Map from Spectra

This function is used to prepare spectral data for hierarchical cluster analysis (HCA) imaging. The HCA imaging routine itself is extensively described in the chapter HCA imaging ('Multivariate Imaging' pull down menu).



The window for data preparation is the same as for other multivariate imaging approaches, such as PCA, KMC or FCM imaging. To start HCA imaging you can either prepare the data from spectral multi file (described here), or load results of a HCA carried out before (see chapter HCA imaging).

Reformatting data sets for HCA imaging: The following data manipulations are performed to prepare data sets for HCA:

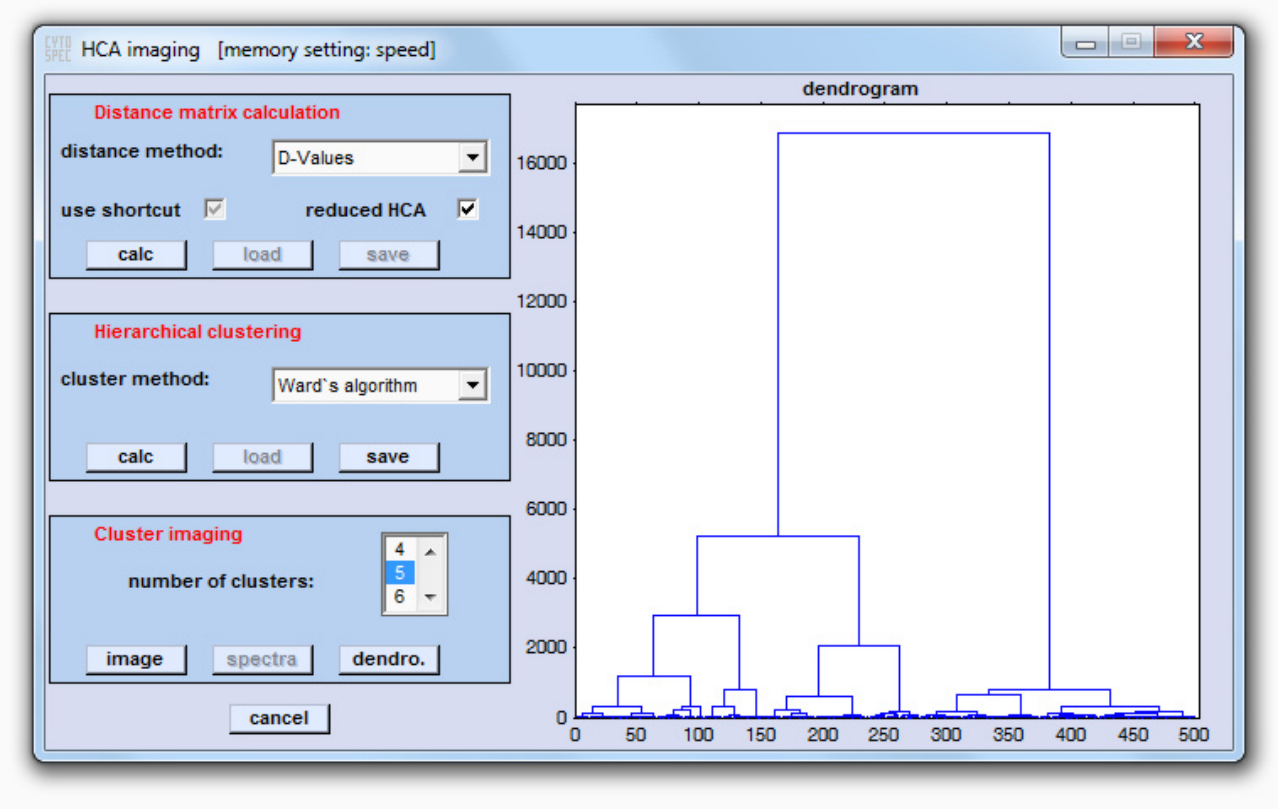
- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test results, or unselected regions of interest, and eliminated.

In the HCA window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may NOT overlap, otherwise an error message will be given. Press 'goto HCA' or 'cancel' by pressing the appropriate button.

Important: Spectra with a negative Quality Test result, or unselected Regions of Interest are automatically excluded from HCA imaging.

Once data preparation has been finished, the HCA imaging dialog box will appear: HCA imaging

Hierarchical Cluster Analysis (HCA): This imaging function performs hierarchical cluster analysis of hyperspectral data and can be used to display the results as HCA images and dendrograms. Spectra with negative ➡ Quality Test results, or unselected ➡ Regions of Interest are excluded from the analysis and appear in HCA images as black pixels. For each spectral class, or cluster, average spectra and standard deviation spectra are calculated which can be stored in an ASCII data format. Furthermore, it is also possible to store, or load, distance data and the results of hierarchical clustering.



The HCA Method:

- First, a distance matrix is calculated which contains information on the similarity of spectra. This matrix is symmetric and of size n x n, where n is the number of spectra. One can choose between five different options to obtain inter-spectral distances.
- Next, the two most similar spectra, that are spectra with the smallest inter-spectral distance, are determined.
- These spectra are combined to form a new object (cluster).
- The spectral distances between all remaining spectra and the new object have to be re-calculated. CytoSpec offers seven different cluster methods.
- A new search for the two most similar objects (spectra or clusters) is initiated. These objects are merged and again, the distance values for the newly formed cluster are determined.
- This procedure is performed n-1 times until only one cluster remains.

How to start cluster imaging?

The HCA imaging function can be either started from the '*Multivariate Imaging --> HCA imaging --> create HCA maps from spectra*' menu or from the pull down menu '*Multivariate Imaging --> HCA imaging --> load HCA data*'. In the latter case one have to load the distance matrix file (extension: *.dis) or the cluster file (extension: *.cls) obtained in earlier program sessions. For details please refer to the chapter ➡ Multivariate Imaging → HCA imaging → create HCA maps from spectra.

Chapters, describing options of the HCA imaging function:

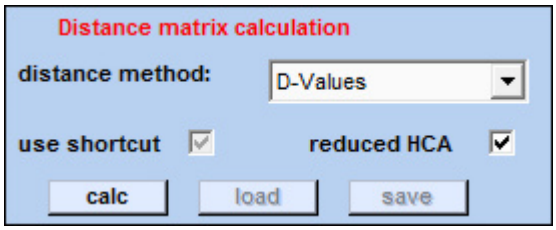
Part I - the distance matrix: Obtaining inter-spectral distances for clustering

Part II - Hierarchical clustering.

Part III - HCA imaging: Reassembling HCA images, obtaining mean cluster spectra.

Part IV - Example of HCA imaging.

Part I - Distance Matrix



Given n objects, a distance or dissimilarity matrix, is a symmetric matrix with zero diagonal elements such that the ij-th element represents how far apart or how dissimilar the i-th and j-th objects are.

calc: starts the calculation of the distance matrix.

load: load distance matrix files (file extensions is '*.dis').

save: save the distance matrix.

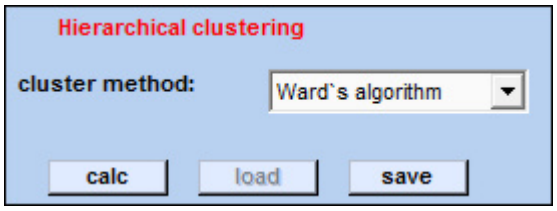
use shortcut: if this option is chosen, the calculation of the distance matrix AND hierarchical clustering are queued that is no further user input will be required. Please select a distance AND a cluster method, even if the distance matrix has not been obtained, before pressing the '*calc*' button. Note also that the distance matrix cannot be stored when this option was selected.

reduced HCA: This HCA imaging option is particularly useful when large datasets are analyzed. It is recommended to choose this option for data files containing more than 128 x 128 spectra. In reduced HCA, the calculation of the distance matrix and hierarchical clustering are carried out on the basis of randomly selected spectra. When finished, mean cluster spectra of the last 50 clusters are obtained. Then, n x 50 distance values between all spectra and mean cluster spectra are obtained (n is the number of pixel spectra in the data set). Cluster memberships are then assigned on the basis of these distance values. Please note, that '*reduced HCA*' is available only for the distance option '*D-values*'.

distance method: pop up menu which allows to select one of the following methods for distance matrix calculation.

1. D-Values	$$r_{jk} = \frac{\left(\sum_{i=1}^{n} x_{ji} \cdot x_{ki}\right) - n \cdot \bar{x}_j \cdot \bar{x}_k}{\sqrt{\left(\sum_{i=1}^{n} x_{ji}^2 - n \cdot \bar{x}_j^2\right) \cdot \left(\sum_{i=1}^{n} x_{ki}^2 - n \cdot \bar{x}_k^2\right)}}$$ (Pearson's correlation coefficient) $$d_{jk} = \left(1 - r_{jk}\right) 1000$$ (D-values)
2. Euclidean distances	$$d_{jk} = \sqrt{\sum_{i=1}^{n} \left(x_{ji} - x_{ki}\right)^2}$$ (Euclidean distance)
3. Normalized Euclidean distances	$$d_{jk} = \sqrt{\sum_{i=1}^{n} \left[\left(\frac{1}{std(x_i)}\right)\left(x_{ji} - x_{ki}\right)\right]^2}$$ (normalized Euclidean distance)
4. Euclidean squared distances	$$d_{jk} = \sum_{i=1}^{n} \left[\left(\frac{1}{std(x_i)}\right)\left(x_{ji} - x_{ki}\right)\right]^2$$ (Euclidean squared distances)
5. City block	$$d_{jk} = \sum_{i=1}^{n} \left(\frac{1}{std(x_i)}\right)\left\lvert x_{ji} - x_{ki}\right\rvert$$ (city block metrik) $$std(x_i) = \sqrt{\frac{1}{(p-1)} \cdot \sum_{m=1}^{p} \left(x_{mi} - \bar{x}_i\right)^2}$$

Part II - Hierarchical clustering



calc: starts hierarchical clustering.

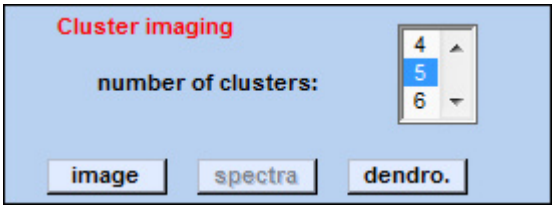
load: load cluster analysis files. The file extension is '*.cls'.

save: save cluster analysis results.

cluster method: here you can select a method for clustering:

1. Average linkage	$d_{i,jk} = \left[\frac{1}{n_i + n_k}\right] \cdot (d_{ji} + d_{ki})$ (average linkage)
2. Single linkage	$d_{i,jk} = \min(d_{ji}, d_{ki})$ (single linkage)
3. Complete linkage	$d_{i,jk} = \max(d_{ji}, d_{ki})$ (complete linkage)
4. Group average	$d_{i,jk} = \left[\frac{n_j}{n_j + n_k}\right] \cdot d_{ji} + \left[\frac{n_k}{n_j + n_k}\right] \cdot d_{ki}$ (group average)
5. Centroid method.	$d_{i,jk} = \left[\frac{n_j}{n_j + n_k}\right] \cdot d_{ji} + \left[\frac{n_k}{n_j + n_k}\right] \cdot d_{ki} - \left[\frac{n_j n_k}{(n_j + n_k)^2}\right] \cdot d_{jk}$ (centroid method)
6. Median algorithm	$d_{i,jk} = \left(\frac{1}{2}\right) \cdot d_{ji} + \left(\frac{1}{2}\right) \cdot d_{ki} - \left(\frac{1}{4}\right) \cdot d_{jk}$ (median method)
7. Wards algorithm.	$d_{i,jk} = \left[\frac{1}{n_i + n_j + n_k}\right] \cdot \left[(n_j + n_i) d_{ji} + (n_k + n_i) d_{ki} - n_i \cdot d_{jk}\right]$ (Wards method)

Part III - HCA imaging

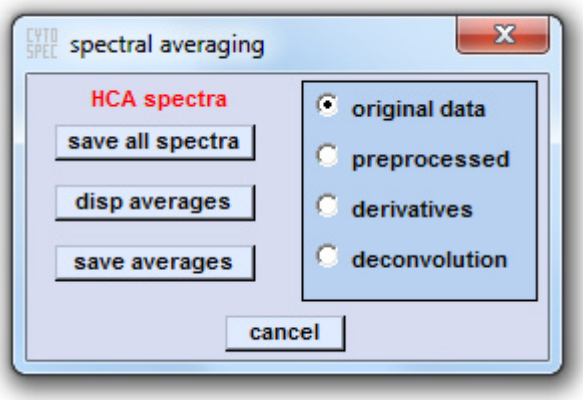


number of clusters: allows selection of the number of clusters used for HCA imaging. Minimum is 2, maximum: 50.

image: displays the cluster image. The number of classes (clusters) are color encoded. The color sequence is determined by the active color map, usually the color map '**ann**' (see function Display Spectra):

dendro: A dendrogram is shown on the display. The dendrogram can be stored as bitmap ('*.bmp') or as an encapsulated postscript ('*.eps') data file. Both functions are available by a activating the context menu of the dendrogram window. Note that dendrograms will show only the last 500 fusion steps.

spectra: this function produces and displays average spectra of each class. After clicking on the '*spectra*' button, a dialog box for choosing the source data block comes up (see screenshot below). This data block is then used for averaging and the creation of the respective standard deviation spectra.

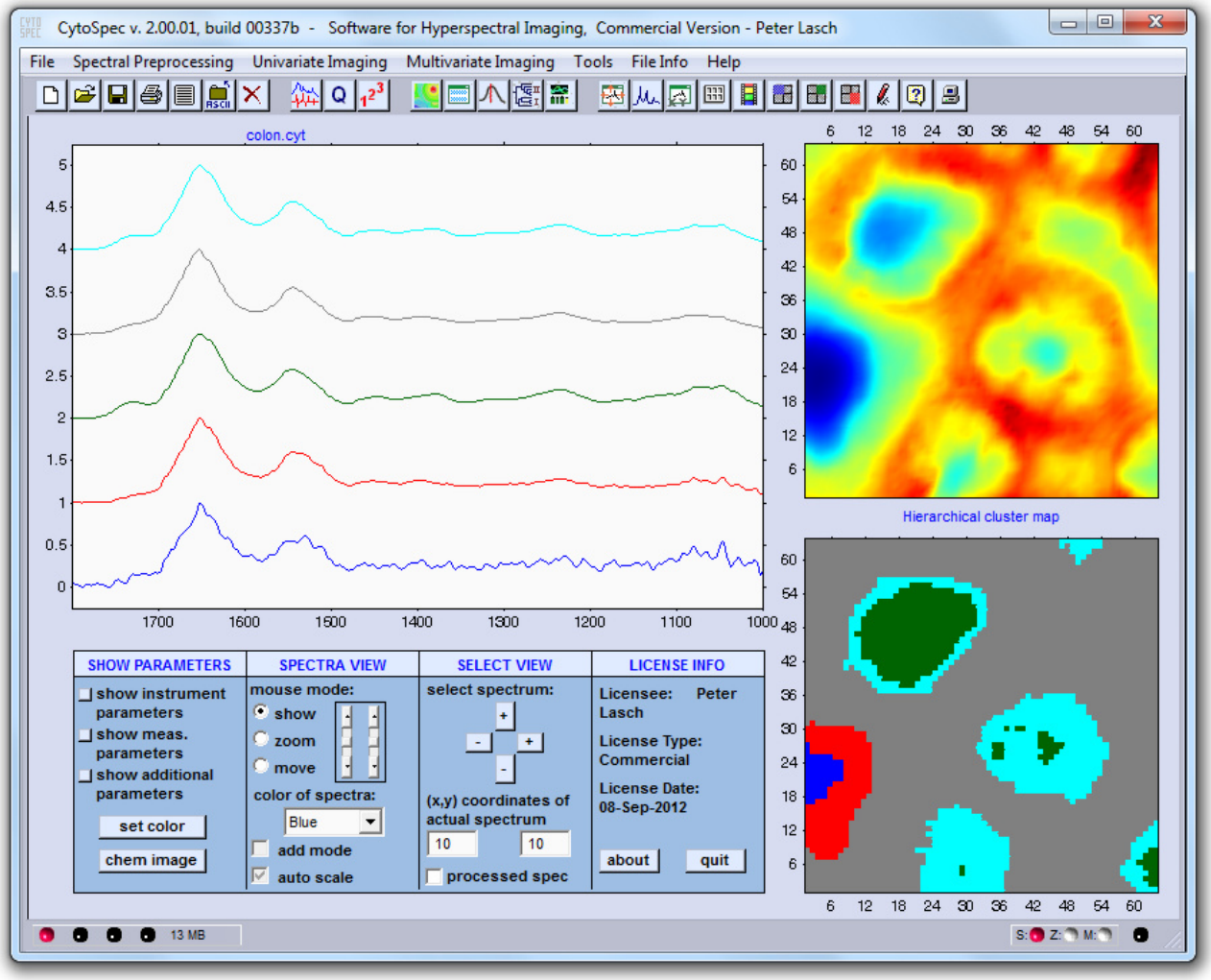


save all spectra: all spectra used for HCA imaging are stored as double column ASCII data. Use the standard windows dialog box to set the path and create the ASCII data files. File names: corename**x_i**: x-th spectrum of the i-th cluster.

disp averages: average spectra of the selected data block are calculated and plotted in the same color like in the cluster map.

save averages: average and standard deviation spectra of the selected data block are calculated. To store these spectra (double column ASCII) use the standard windows dialog box to set the path and chose a file name. Average spectra and standard deviation spectra of the i-th cluster are stored per default in separate files (corename_**i** - for average spectra and corename_**std_i** for the standard deviation spectra).

Part IV - Example of HCA imaging:



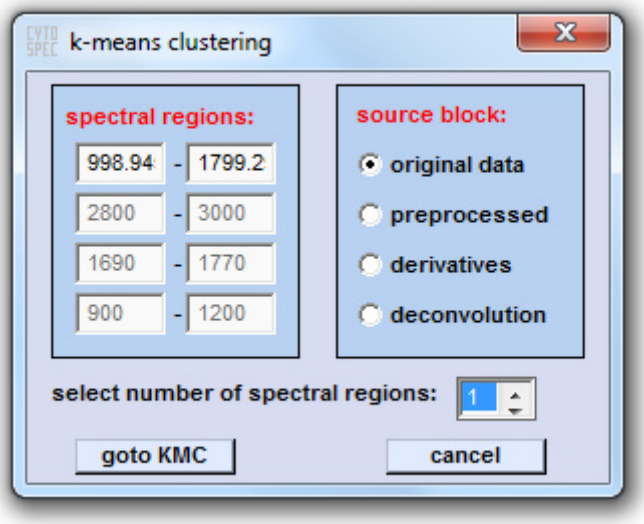
This example shows the results of HCA imaging (5 classes). Spectral data were acquired by the use of a 64 x 64 mid-infrared MCT focal plane array detector. For HCA imaging, Ward's algorithm was used. Spectral distances were computed as D-values. The panel to the left displays Min-Max Normalized average spectra, which were encoded by the same color utilized for displaying the cluster in the HCA image (HCA was carried out on the basis of the file 'colon.cyt'. which can be found in the /testdata/bin/CytoSpec/ directory).

Reference to the literature:

- Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta.* **2004** 1688(2):176-86.
- Lasch P, Diem M, Hänsch W & Naumann D. Artificial neural networks as supervised techniques for FT-IR microspectroscopic imaging. *Journal of Chemometrics* **2007** Vol. 20(5):209-220.

KMC Image Segmentation - k-Means Cluster Imaging

This function is used to prepare spectral data for k-means cluster (KMC) imaging.



The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, PCA, or FCM imaging. Before KMC imaging can be started you have to prepare the data from a spectral multi file.

Reformatting data sets for KMC imaging: The following data manipulations are performed to prepare data sets for KMC:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test result, or unselected regions of interest, and eliminated.

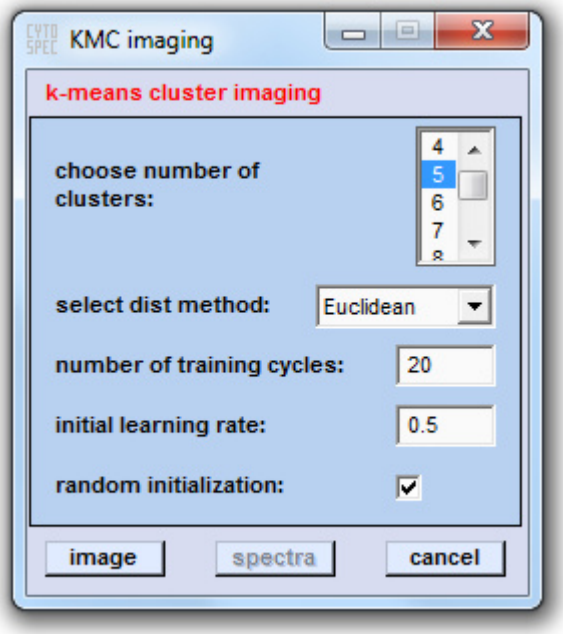
In the KMC window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may overlap, that is they may appear two-, three-, or even fourfold in the reformatted spectra. Press '*goto KMC*' or '*cancel*' by pressing the appropriate button.

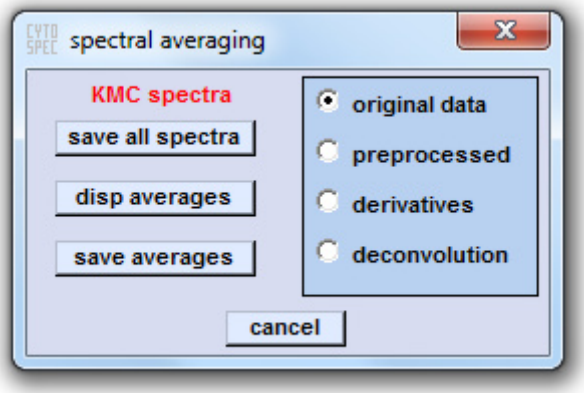
Important: Spectra with a negative ➡ Quality Test result, or unselected ➡ Regions of Interest are automatically excluded from KMC imaging.

Once data preparation has been finished, the following KMC imaging dialog box will appear:

How to start KMC imaging? Once data preparation has been finished, the KMC imaging dialog box appears. Indicate how many classes are assumed to be present in the data set ('*number of cluster*') and the number of training cycles (default: 100). Press the '*image*' button if you wish to start the application or hit '*cancel*' to abort the function

Button '*spectra*': This option becomes available when the k-means clustering calculations are finished. When this button is pressed, a new window comes up that offers additional options such as the calculation of cluster mean spectra or sorting individual map spectra by its cluster membership





save all spectra: all spectra used for KMC imaging are stored as double column ASCII data. Use the standard windows dialog box to set the path and create the ASCII data files. File names: corename**x_i**: x-th spectrum of the i-th cluster.

disp averages: average spectra of the selected data block are calculated and plotted in the same color like in the cluster map.

save averages: average and standard deviation spectra of the selected data block are calculated. To store these spectra (double column ASCII) use the standard windows dialog box to set the path and chose a file name. Average spectra and standard deviation spectra of the i-th cluster are stored per default in separate files (corename_**i** - for average spectra and corename_**std**_i for the standard deviation spectra).

Principles of k-means clustering: The algorithm of k-means clustering has been suggested by J.B. MacQueen in 1967:
J.B. MacQueen. In L.M. LeCam and J. Neymann (eds) *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967 281-297.
Related web links (Wikipedia):

- k-means clustering
- k-means algorithm

In the CytoSpec implementation, MacQueens k-means cluster algorithm is used. k-means clustering is a non-hierarchical clustering method, which obtains a "hard" (crisp) class membership for each spectrum, that is the class membership of an individual spectrum can be take only the values of zero or one. It uses an iterative algorithm to update randomly selected initial cluster centers, and to obtain the class membership for each spectrum, assuming well-defined boundaries between the clusters. MacQueens

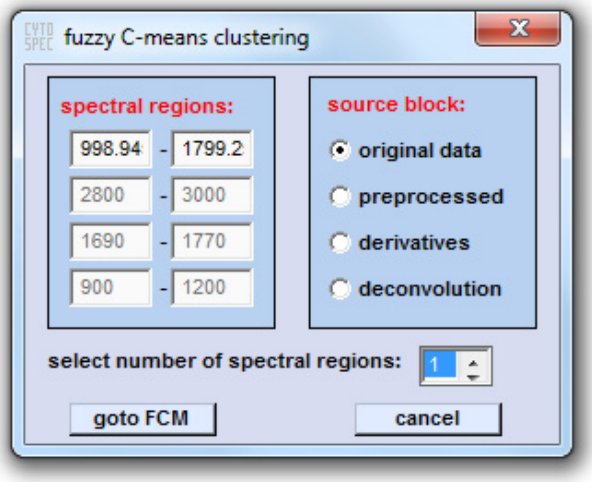
iterative algorithm of KMC can be described as follows: Spectra are illustrated as points in a p-dimensional space (p is the number of features of the spectra. In this space a number of k points is initially chosen, where each point represents a cluster to be made. Then, distance values between the points and all objects (spectra) are calculated. Objects are assigned to a cluster on the basis of a minimal distance value. Next, centroids of the clusters are calculated and distance values between the centroids and each of the objects are re-calculated. Then, if the closest centroid is not associated with the cluster to which the object currently belongs, the object will switch its cluster membership to the cluster with the closest centroid. The centroid's positions are re-calculated every time a component has changed the cluster membership. This continues until none of the objects has been re-assigned.

Reference to the literature:

- Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta.* **2004** 1688(2):176-86.

FCM Cluster Image Segmentation - Fuzzy C-Means Cluster Imaging

This function is used to prepare spectral data for fuzzy C-means cluster (FCM) imaging.



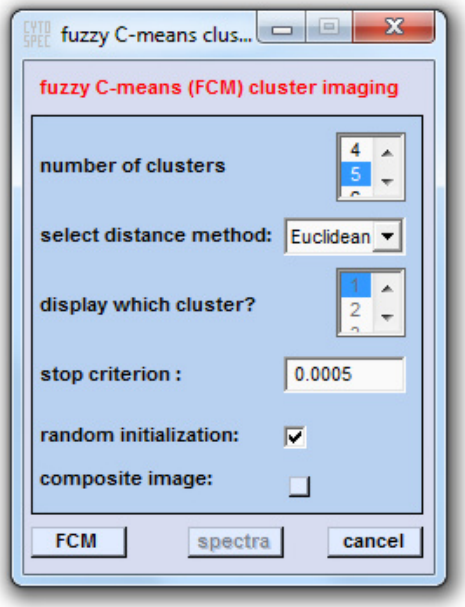
The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, PCA, or KMC imaging. Before FCM imaging can be started you have to prepare the data from a spectral multi file..

Reformatting data sets for FCM imaging: The following data manipulations are performed to prepare data sets for FCM:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test result, or unselected regions of interest, and eliminated.

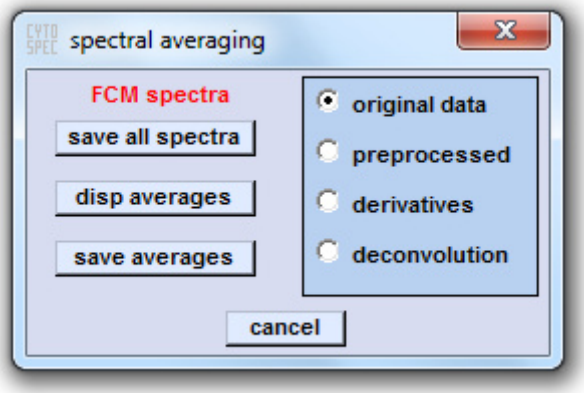
In the FCM window, select the number of spectral regions (min:1; max:4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may overlap, that is they may appear two-, three-, or even fourfold in the reformatted spectra. Press '*goto FCM*' or '*cancel*' by pressing the appropriate button.

Important: Spectra with a negative ➡ Quality Test result, or unselected ➡ Regions of Interest are automatically excluded from FCM imaging.



How to start FCM imaging? Once data preparation has been finished, the FCM imaging dialog box appears. Indicate how many classes are assumed to be present in the data set ('how many clusters to search for') and an exit criterion ('minimum amount of improvements', default: 0.0005). Press the 'C-means' button if you wish to start the application or hit 'cancel' to abort the function. When the calculations are finished select a cluster that should be used to reassemble the image ('display which cluster'). Note that images can be reassembled from one cluster, only**!**

Button '*spectra*': This option becomes available when the fuzzy C-means clustering calculations are finished. When this button is pressed, a new window comes up that offers additional options such as the calculation of cluster mean spectra or sorting individual map spectra by its cluster membership. By pressing any of the buttons (except button 'done') a new FCM imaging approach is initiated by using the settings of the FCM imaging dialog box.



save all spectra: all spectra used for FCM imaging are stored as double column ASCII data. Use the standard windows dialog box to set the path and create the ASCII data files. File names: corename**x**_**i**: x-th spectrum of the i-th cluster.

disp averages: average spectra of the selected data block are calculated and plotted.

save averages: average and standard deviation spectra of the selected data block are calculated. To store these spectra (double column ASCII) use the standard windows dialog box to set the path and chose a file name. Average spectra and standard deviation spectra of the i-th cluster are stored per default in separate files (corename_**i** - for average spectra and corename_**std**_i for the standard deviation spectra).

Principles of fuzzy C-means clustering: Principles of FCM clustering are given in the following publication:

- J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*, **1981** New York. Plenum Press.

Related web links:

- fuzzy C-means clustering (Wikipedia)

FCM clustering is a non-hierarchical clustering method. This clustering technique partitions objects into groups (cluster) whose members show a certain degree of similarity. Unlike k-means clustering, the output of FCM clustering is a membership function, which defines the degree of membership of a given spectrum to the clusters. The values of the membership function can vary between one (highest degree of cluster membership) and zero (no class membership), where the sum of the C cluster membership values for one object equals one.
Thus, this method departs from the classical two-valued (0 or 1) logic, and uses "soft" linguistic system variables and a continuous range of true values in the interval [0,1]. FCM imaging uses a fuzzy iterative algorithm to calculate the class membership grade for each spectrum. The iterations in FCM clustering are based on minimizing an objective function, which represents the distance from any given data point (spectrum) to the actual cluster center weighted by that data points membership grade.
The advantage of the fuzzy C-means clustering over k-means clustering is that both outliers and data, which display properties of more than one class can be characterized by assigning nonzero class membership values to several clusters. In the similarity maps assembled by FCM clustering the

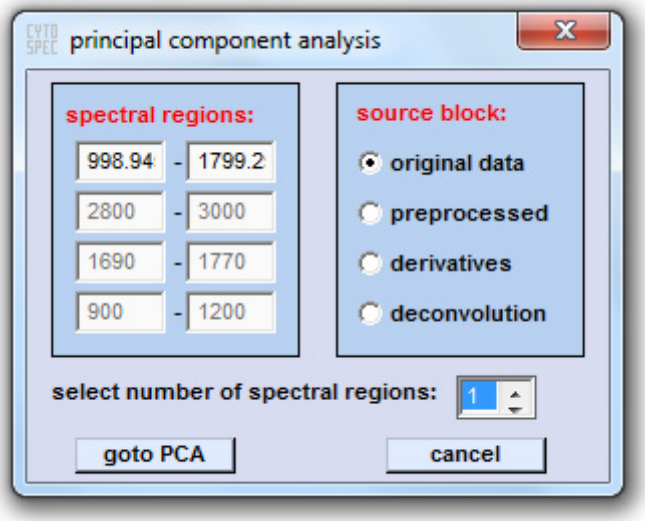
membership values are encoded by the colormap, that is by the color intensities in the case of single color colormaps.

Reference to the literature:

- Lasch P, Haensch W, Naumann D, Diem M. Imaging of colorectal adenocarcinoma using FT-IR microspectroscopy and cluster analysis. *Biochim Biophys Acta.* **2004** 1688(2):176-86.

Create PCA Maps from Spectra

This function is used to prepare spectral data for principal component analysis (PCA) imaging. The PCA imaging routine itself is extensively described in the chapter ➡ PCA imaging ('Multivariate Imaging' pull down menu).



The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, KMC or FCM imaging. To start PCA imaging you can either prepare the data from spectral multi file (described here), or load results of a PCA carried out before (see chapter ➡ PCA imaging).

Reformatting data sets for PCA imaging: The following data manipulations are performed to prepare data sets for PCA:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test results, or unselected regions of interest, and eliminated.

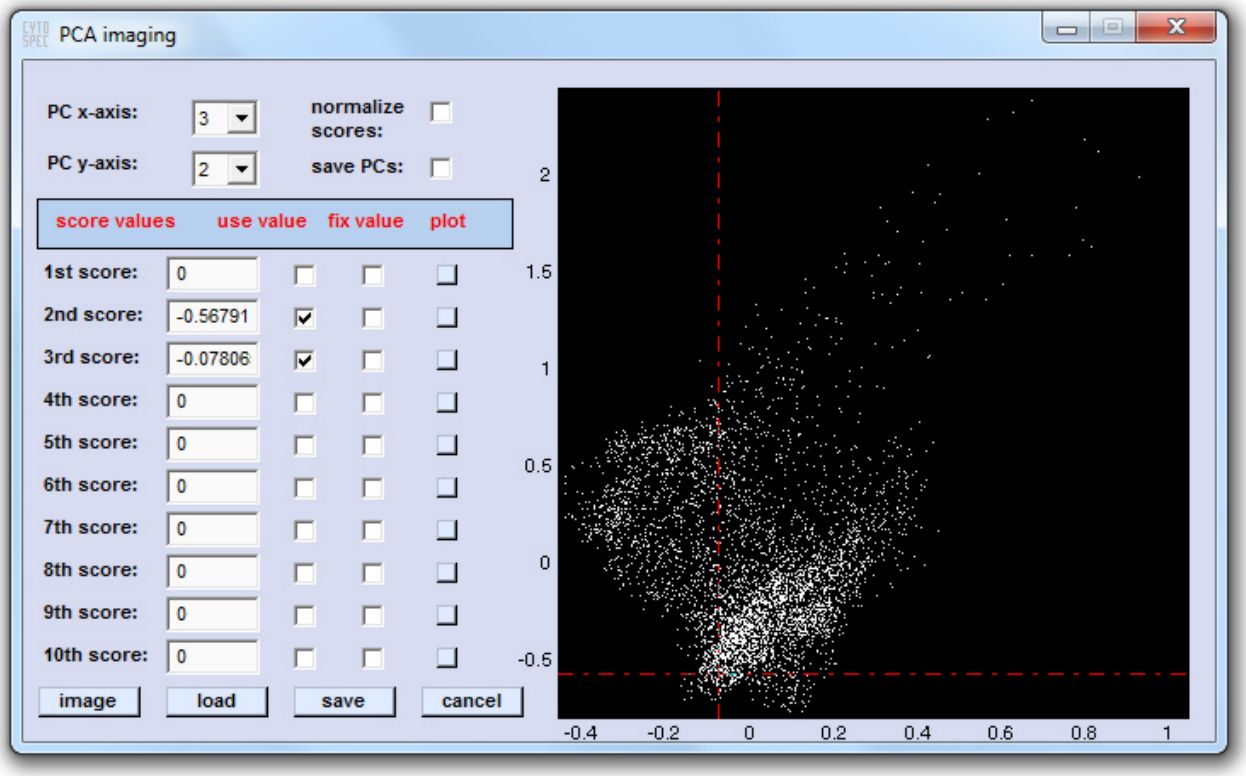
In the PCA window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may NOT overlap, otherwise an error message will be given. Press 'goto PCA' or 'cancel' by pressing the appropriate button.

Important: Spectra with a negative ➡ Quality Test result, or unselected ➡ Regions of Interest are automatically excluded from PCA imaging.

Once data preparation has been finished, the PCA imaging dialog box will appear: ➡ PCA imaging

PCA Imaging - Imaging Using Principal Component Analysis

Principles of PCA: PCA is a linear transformation in which the (spectral) data are transferred into a new coordinate system. In this new coordinate system, the largest data variance points to the direction of the first coordinate, which is also called the first principal component (pc), the second largest variance on the second pc, and so forth.
PCA is therefore a transformation that re-arranges the data according to the data's intrinsic variance: most of the variance is contained in the lower-order principal components while higher-order pc's are supposed to contain mainly noise. Reduction of dimensionality by PCA can be effectively achieved by omitting higher-order principal components.



How to start PCA imaging: One can start PCA imaging either from the '*Image Manipulations --> PCA maps*' pull down menu or from the menu '*Multivariate Statistics --> PC analysis (PCA)*'. In the first case one have to load a PCA file (*.pca) obtained in earlier program sessions. Please refer to the chapter ➡ Multivariate Statistics → PCA maps if you want to produce PCA images directly from a hyperspectral data set.

Define the number of dimensions: to specify which PCs should be used for imaging, check the appropriate checkboxes of the column '*use value*'. In the example above the principal components one and two are activated. The CytoSpec program permits the use of the first 10 principal components.

Definition of individual score coefficients: first, select the principal components to be displayed from the popup menus in the upper part of the PCA imaging window (PC x- or y-axis). Then, click into the score plot window to the right. The respective coordinates (mass centers) of this action are transferred to the edit fields indicated as 1st to 10th score. Alternatively, it is possible to manually type the respective coordinates into these boxes.

Normalization of the score coefficients: checking the checkbox '*normalize scores*' causes normalization of distances between score coefficients and 'mass centers' such that the maximum distance for all principal components equals one.

What does 'fix value' mean? This option permits to fix the coordinates of a given mass center in the n-th dimension, irrespective of mouse manipulations in the plot to the right. This option may be useful

when searching for an optimal contrast of a PCA image.

save PCs: if this option is chosen, the first 10 principal components are stored (see description of the button '*save*' below for details).

imaging: using the actual coordinates of the mass centers, PCA images are plotted into the lower right panel of the main window.

load: opens a standard window for opening files of the format *.pca.

save: allows to save score coefficients and principal components. File extension will be *.pca. If the checkbox '*save PC's*' was checked, the first 10 principal components are stored as separate double column ASCII files. These files are stored in the same directory as the *.pca-file.

cancel: closes the PCA imaging window. Data not stored are lost.

Note that **s**pectra with negative ➡ Quality Test results, or unselected ➡ Regions of Interest are excluded from the analysis and appear in PCA images as black pixels

Reference to the literature:

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

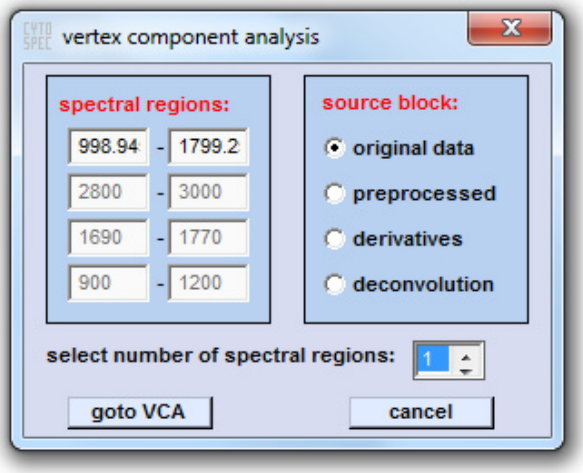
VCA Imaging - Imaging Using Vertex Component Analysis

Principles of VCA: VCA (vertex component analysis) is an unsupervised method to rapidly unmix hyperspectral data. The algorithm was initially developed by J. Nascimento and J. Dias. The idea of VCA can be summarized as follows: Given a set of mixed spectral (multispectral or hyperspectral) vectors, linear spectral mixture analysis, or linear unmixing, aims at estimating the number of reference substances, also called endmembers, their spectral signatures, and their abundance fractions. Unsupervised endmember extraction by VCA exploits the following facts: the endmembers are the vertices of a simplex and the affine transformation of a simplex is also a simplex. Furthermore, VCA assumes the presence of pure pixels in the data. The algorithm iteratively projects data onto a direction orthogonal to the subspace spanned by the endmembers already determined. The new endmember signature corresponds to the extreme of the projection. The algorithm iterates until all endmembers are found.

Related links:

- J. Nascimento and J. Dias, "Vertex Component Analysis: A fast algorithm to unmix hyperspectral data", *IEEE Transactions on Geoscience and Remote Sensing* **2005** vol. 43, no. 4, pp. 898-910
- Unsupervised unmixing of hyperspectral imagery using the constrained positive matrix factorization by Yahya M. Masalmah. July **2007**. Dissertation, University of Puerto Rico. Chair: Miguel Vel´ez-Reyes, Major Department: Computing and Information Science and Engineering

To start VCA imaging one have to prepare the spectral data. In order to do so please select '*VCA imaging*' from the '*multivariate imaging*' pulldown menu.



The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, KMC or FCM imaging.

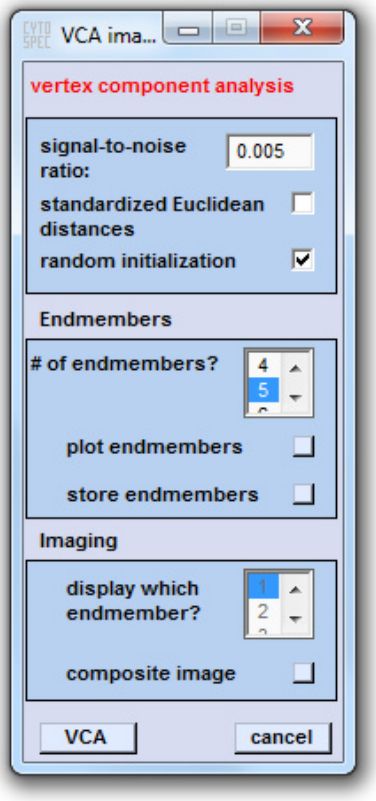
Reformatting data sets for VCA imaging: The following data manipulations are performed to prepare data sets for VCA:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test results, or unselected regions of interest, and eliminated.

In the VCA window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may NOT overlap, otherwise an error message will be given. Press '*goto VCA*' or '*cancel*' by pressing the appropriate button.

Important: Spectra with a negative Quality Test result, or unselected Regions of Interest are automatically excluded from VCA imaging.

Once data preparation has been finished, the following VCA imaging dialog box will appear:



signal-to-noise ratio: an exit criterion

standardized Euclidean distances: this checkbox permits to choose whether VCA is carried out in Euclidean or normalized Euclidean space

number of endmembers: please indicate the number of endmembers

plot endmembers: spectral signatures of the endmembers are plotted into the spectral window of the main gui

store endmembers: allows to store endmember spectra as double collumn ASCII data

display which endmember: choose a endmember which abundance fraction is used for VCA imaging

composite image: plots a composite image using abundance fractions of all endmembers

VCA: starts the VCA routine

cancel: press cancel to exit. Data not stored are lost

Application of VCA in vibrational spectroscopic imaging:

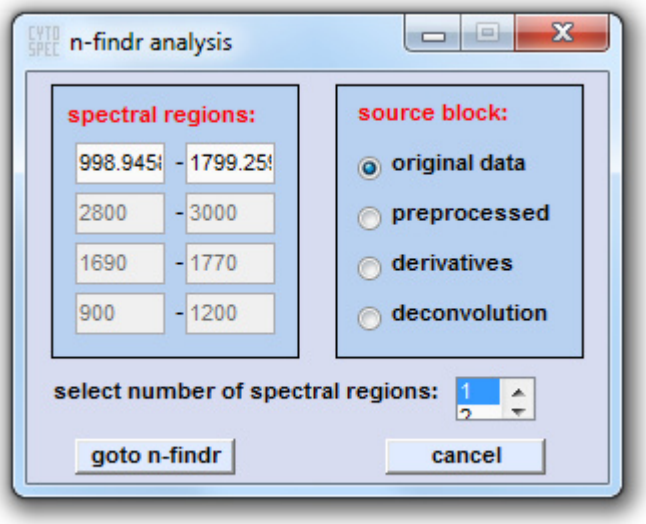
- Chernenko T, Matthäus C, Milane L, Quintero L, Amiji M, Diem M. Label-free Raman spectral imaging of intracellular delivery and degradation of polymeric nanoparticle systems. *ACS Nano*. **2009**. 3(11):3552-9.

Principles of n-findr endmember extraction: Use this link for details (Winter, Michael E., Pacific Spectral Technology).

Related links:

- M. E. Winter, M. R. Descour, and S. S. Shen. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. volume 3753, pages 266-275, Denver, CO, USA, October **1999**. *SPIE*.
- Winter, Michael E., "Fast Autonomous Spectral End-member Determination In Hyperspectral Data", *Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing*, Vol. II, pp 337-344, Vancouver, B.C., Canada, **1999**.

To start n-findr imaging one have to prepare the spectral data. In order to do so please select '*n-findr imaging*' from the '*multivariate imaging*' pulldown menu.



The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, KMC or VCA imaging.

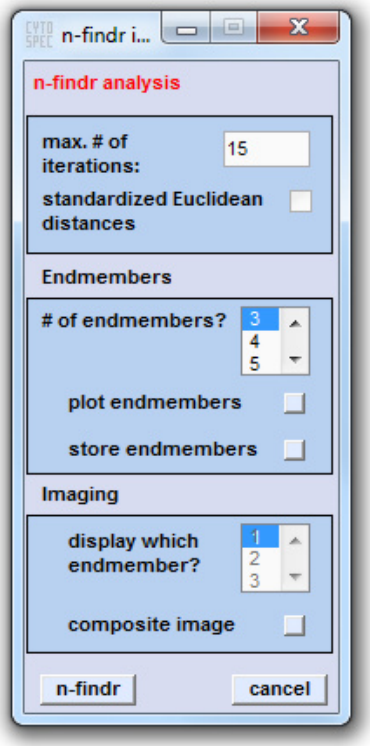
Reformatting data sets for n-findr imaging: The following data manipulations are performed to prepare data sets for n-findr endmember extraction:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test results, or unselected regions of interest, and eliminated.

In the n-findr window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may NOT overlap, otherwise an error message will be given. Press '*goto -n-findr*' or '*cancel*' by pressing the appropriate button.

Important: Spectra with a negative Quality Test result, or unselected Regions of Interest are automatically excluded from n-findr endmember extraction imaging.

Once data preparation has been finished, the following n-findr imaging dialog box will appear:



max number of iterations: the exit criterion of the n-findr endmember extraction method.

standardized Euclidean distances: inactive

number of endmembers: please indicate the number of spectral endmembers

plot endmembers: spectral signatures of the endmembers are plotted into the spectral window of the main gui

store endmembers: allows to store endmember spectra as double collumn ASCII data

display which endmember: choose a endmember which abundar fraction is used for n-findr imaging.

composite image: plots a composite image using abundance fractions of all endmembers

n-findr: starts the n-findr routine

cancel: press cancel to exit. Data not stored are lost

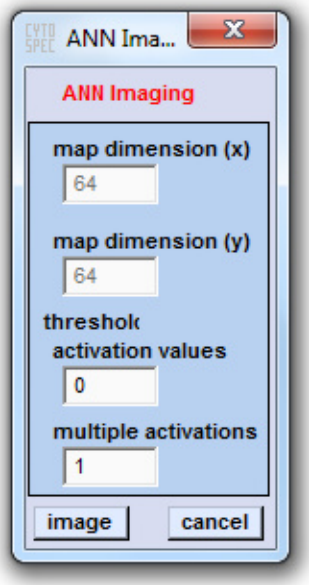
Application of n-findr imaging in vibrational hyperspectral imaging:

- Lau K, Hedegaard MA, Kloepper JE, Paus R, Wood BR, Deckert V. Visualization and characterisation of defined hair follicle compartments by Fourier transform infrared (FTIR) imaging without labelling. *J Dermatol Sci*. **2011** 63(3):191-8.

- Bergner N, Krafft C, Geiger KD, Kirsch M, Schackert G, Popp J. Unsupervised unmixing of Raman microspectroscopic images for morphochemical analysis of non-dried brain tumor specimens. *Anal Bioanal Chem*. **2012** 403(3):719-25.

This function of the CytoSpec program is designed to re-assemble hyperspectral images on the basis of classification results of artificial neural networks (ANN). In this function, result files (*.res) of the **Stuttgart Neural Network Simulator** (SNNS) are analyzed and directly converted into false-colored ANN maps. Furthermore, checks for activation thresholds and multiple activations can be carried out.

The Network Simulator was developed at the "Institut für Parallele und Verteilte Höchstleistungsrechner" (IPRV) of the Universität Stuttgart (Germany). The SNNS can downloaded for free at ➡ ftp://ftp.informatik.uni-stuttgart.de.



xdim in ANN map: the number of measurement points in x-direction.

ydim in ANN map: the number of measurement points in y-direction.

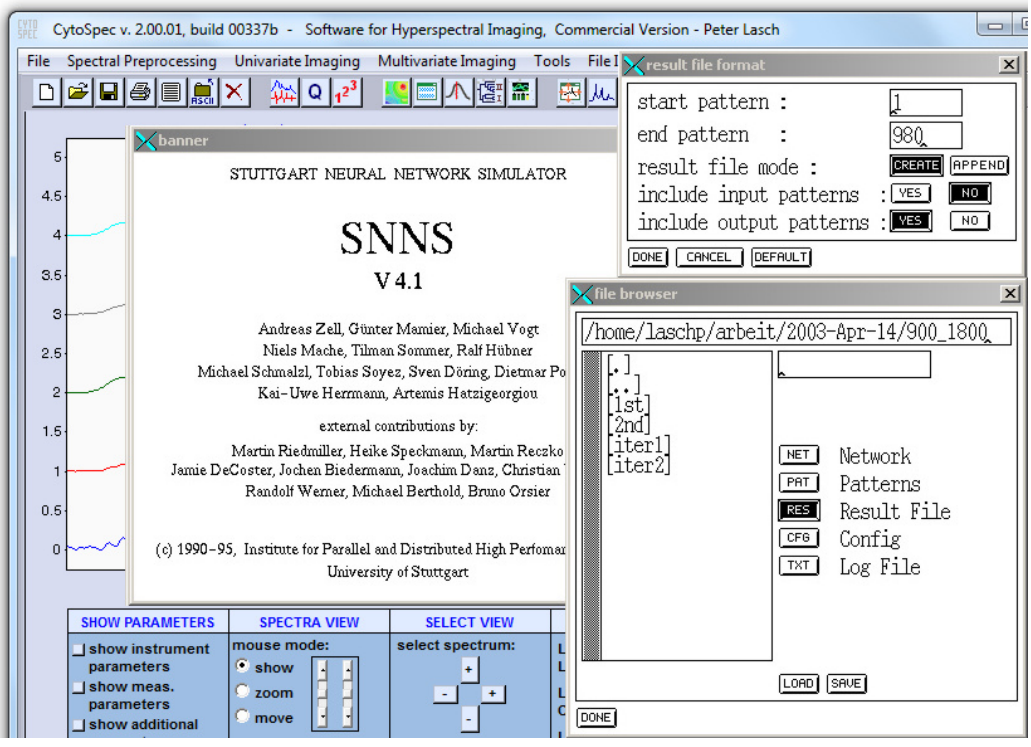
threshold for activations: threshold which defines the minimum allowed output activation. Activations below this threshold are set to zero and the black color is assigned to the corresponding pixels. Set this threshold to 0 if you want to omit this test.

threshold for multiple activations: threshold which defines the maximum allowed activation of the second-highest activated neuron. If a spectrum meets this criterion, it is tested as negative and again the black color is assigned to corresponding pixel. One can omit this test by setting the threshold to 1.

image: opens the standard windows file browser. Please select a path and a valid SNNS result file.

cancel: aborts the ANN imaging function

IMPORTANT: Please check carefully the sequence of the input patterns (spectra) when compiling the SNNS pattern file. Note that only the pattern sequence will define the spatial (x,y) positions of individual spectra. Select the option 'include output patterns' and unselect the option 'include input patterns' when saving '*.res' files (see screenshot of the 'result file format' window below).



Format of SNNS result files:

```
SNNS result file V1.4-3D
generated at Tue Mar 25 11:24:47 1997

No. of patterns      : 980
No. of input units   : 76
No. of output units  : 4
startpattern         : 1
endpattern           : 980
teaching output included
#1.1
1 0 0 0
0.06227 0.82028 0.01888 0.00005
#2.1
1 0 0 0
0.97587 0.35621 0.00001 0.00089
#3.1
1 0 0 0
0.99435 0.28643 0.00001 0.00046
  .....

#979.1
1 0 0 0
0 0.05502 0.98514 0.21558
#980.1
1 0 0 0
0 0.14128 0.81746 0.61632
```

In the example given above, 980 spectra obtained from a rectangular area (20 x 19 spectra) were analyzed. The number of pre-defined classes in the teaching phase of ANN model development was four.
The first line of the first pattern (#1.1) shows the *a priori* class assignment (target pattern), while the second line displays the ANN test results for this particular spectrum. The maximum activation was found for the second output neuron (0.82028), indicating a *posteriori* class assignment of this individual spectrum to class # two. CytoSpec automatically analyzes the *posteriori* assignments for all spectral sub-pattern contained in the *.res file and assigns specific colors to each class. Images are produced by combining colors with spatial (pixel) positions of the spectra assuming rectangular regions and equal distances between pixels in x- and y-direction, respectively.

The following types of ANNs were tested to be compatible:

- multilayer perceptron (MLP) networks consisting of three layers of neurons (input layer, hidden layer, output layer)
- feed-forward propagation of activations, shortcut connections are allowed
- teaching functions: backpropagation, resilient backpropagation (rprop), quickpropagation (quickprop)

CytoSpec also permits basic tests for multiple activations (i.e. if the second-highest activation is larger than a defined threshold) and for a required minima of activation. If one of these tests is negative the black color is assigned to the corresponding pixel. In the command line window (and the log-file) you can find additional information on the test results.

Reference to the literature:

- Lasch P, Haensch W, Kidder L, Lewis EN. Naumann D. [2002] Colorectal Adenocarcinoma Characterization by Spatially Resolved FT-IR Microspectroscopy. *Appl. Spectrosc.* **2002** 56 (1). 1-9

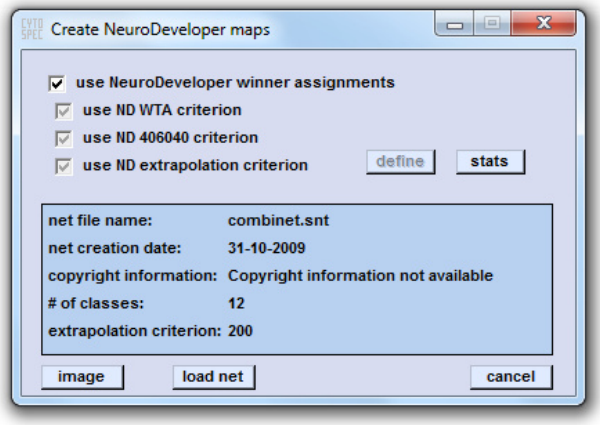
- Lasch, P. & Naumann, D. FT-IR Microspectroscopic Imaging of Human Carcinoma Thin Sections Based on Pattern Recognition Techniques. *Cellular and Molecular Biology* **1998** 44(1). pp. 189-202

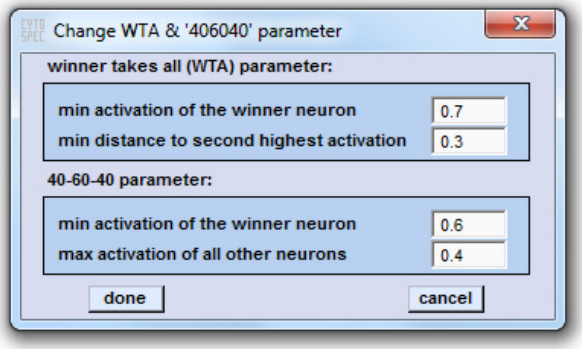
Image Segmentation with Synthon's NeuroDeveloper(TM) Neural Network Simulator

The function called '*Synthon maps*' is basically an interface between CytoSpec and Synthon's NeuroDeveloper (TM), a software for teaching and validating artificial neural network models with spectra from various origins (e.g. IR, Raman, MS spectra). Based on neural network models, the interface can be used to re-assemble ANN images from CytoSpec's original data set. Spectral pre-processing, features selection and ANN classification of *a priori* unknown spatially resolved IR maps can be easily performed in one step. This is achieved by utilizing the runtime environment of the NeuroDeveloper software which does not require a software license from Synthon. Spectral data can be therefore classified without the NeuroDeveloper software on the basis of predefined network libraries. The NeuroDeveloper is, however, required if you wish to create and validate own neural network models.

Synthon GmbH, contact address: Analytics and Pattern Recognition, Im Neuenheimer Feld, 69120 Heidelberg, GERMANY
phone: +49 6221 50 257 900
fax: +49 6221 50 257 909
email: info@synthon-analytics.com
internet: http://www.synthon-analytics.de

To start the function select '*Synthon maps*' from the '*image manipulation*' pull down menu:





load allows to browse the directory structure and load the NeuroDeveloper network library (*.snt) file. After loading the '*image*' and '*stats*' buttons are activated.

image displays the ANN map in CytoSpec's main window

stats gives an overview on ANN classification statistics (see screenshot below)

use NeuroDeveloper winner assignments: if this checkbox is checked, CytoSpec uses the NeuroDeveloper settings for ANN classification. The NeuroDeveloper settings can be modified by unselecting this checkbox.

use ND WTA criterion: the NeuroDeveloper settings for the Winner Takes All (WTA) criterion are used. You can modify these values by deselecting this checkbox (see button '*define*' and chapter below)

use ND 406040 criterion: the NeuroDeveloper settings for the 40/60/40 (or 406040) criterion are used. You can modify these values by deselecting this checkbox (see button '*define*' and chapter below)

use ND extrapolation criterion: the NeuroDeveloper settings for the extrapolation are used. This option can be deactivated.

define opens a window for entering own WTA and 406040 criteria (see screenshot to the left).

cancel aborts the '*Synthon maps*' function

Spectra that have failed the tests appear within the ANN maps as black pixels.
The evaluation of the activations calculated by the network is performed with the analysis functions WTA and 40-20-40. For this evaluation the scores and the distribution of the activations from the output neurons are taken into account.

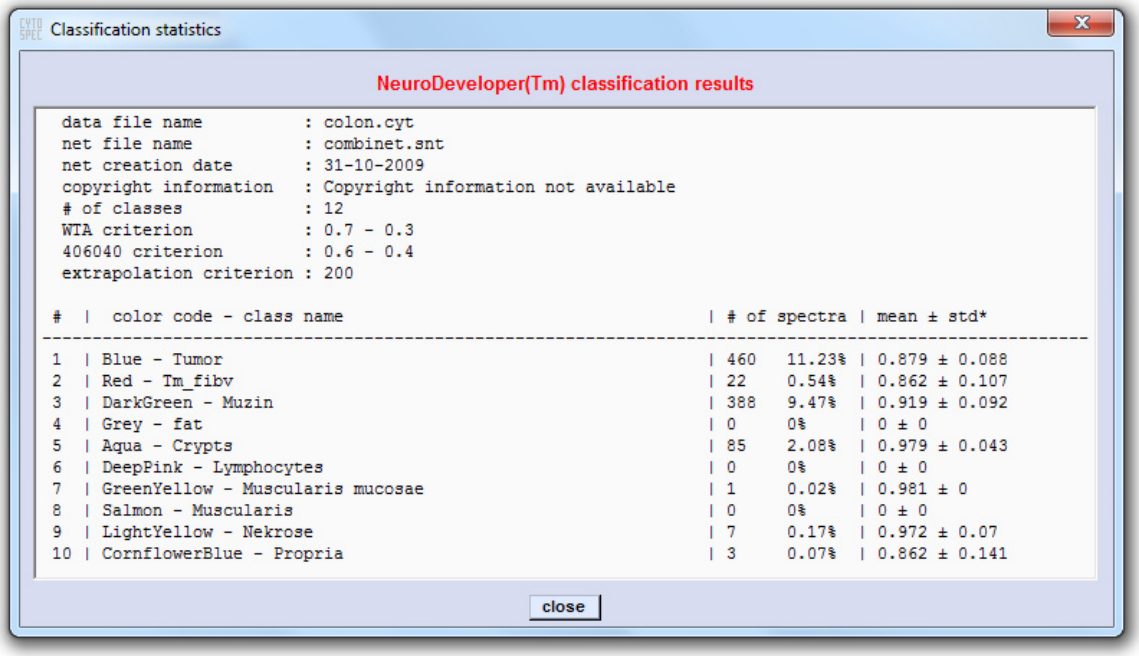
WTA criteria:
WTA stands for winner takes all, which means the classification depends on the highest output activation. A spectrum will only be classified if its output is greater than the defined minimum activation of winner neuron (default 0.7) and the minimum distance to next activation (default 0.3). Otherwise the classification will not be considered correctly and the spectrum remains unclassified.

'406040' criteria
The 40-20-40 function works differently. The activation of one neuron has to exceed 0.6 (default, above 60 percent of the activation range). All other activations of further classes have to be below 0.4 (below 40 percent of the activation range). Otherwise, the pattern remains unclassified.

'extrapolation' criterion
A general problem with different classification methodologies is the potential misclassification due to undesired or unexpected extrapolation. This occurs, when the training and validation datasets do not comprise all classes or the entire range of a feature needed for a given classification problem. In this case, any classification method, including ANNs, would not be representative for the given problem. Data of this type should rather be termed not classified. The NeuroDeveloper uses a distance value derived from the training and validation dataset, to determine an extrapolation problem. The maximum distance of a pattern to its corresponding class is calculated and set to 100. During the classification of a new pattern by the ANN, the distance of the new pattern is calculated and set into relation. In case the calculated extrapolation value of the class, identified by the neural network, exceeds 100, an extrapolation occurs. The default value to determine patterns as unclassified is proposed to be set to 200. The value should be set greater than 100, the smaller the value, the stricter the threshold.
Please note: if the checkbox '*use NeuroDeveloper winner assignments*' was checked, CytoSpec does not perform an analysis of the WTA, 406040 and extrapolation results. The NeuroDeveloper software excludes spectra from further analysis in the following way:

1. either both, the WTA, or the 406040 criteria failed.
2. failed classification based on the extrapolation criterion

The definition of '*failed classifications*' in CytoSpec is different. CytoSpec defines spectra as unclassified if an individual spectrum failed one of the three criteria. For this reasons, the classification statistics may depend on the program used for analysis.



Screenshot of the NeuroDeveloper(TM) classification statistics window

A. Compilation of data sets for teaching and internal validation

- Load an hyperspectral data set and produce a hyperspectral map (e.g. chemical map, HCA map)
- Obtain the context menu of the maps by clicking with the right mouse button over the spectral map.
- Choose '*class 1*' --> and '*start*' if you want to assign spectra to class 1. Now, you are in the '*select spectra*' mode. In this mode, the mouse cursor changes its appearance (arrow plus cross).
- You can select now an unlimited number of spectra by left mouse clicks (in this mode; spectra will be not displayed). The spatial coordinates will be given in the command line window.
- To stop the selection mode, choose '*selection mode off*' from the context menu. Alternatively, you can immediately start to assign spectra to class 2 by selecting '*class 2*' --> and '*start*' from the context menu. In this way, spectra can be assigned to up to 10 distinct classes.
- If all spectra are selected, stop the selection mode by '*selection mode off*'. In the normal '*show spectra*' mode the mouse pointer will regain its normal appearance (arrow).
- In order to export spectra select the '*export*' --> '*x,y ASCII*' from the file pull down menu. A window with the title '*convert into a x,y-ASCII data format*' appears. Check the checkbox '*export selection*'. Please use the default settings for all other options. Make sure, that the data block of original absorbance spectra is exported.
- Press button '*export*' and store the spectra in a folder of your choice. Spectra of class 1 can be identified by the extension '*_1.dat', spectra by class 2 are named '*_2.dat' and so forth.
- Steps 1-8 should be repeated for a number of maps. It is recommended to use consistent file names to assure consistent class assignments.
- Split the spectral data into a subset for teaching (ca. 65 % of the spectra) and internal validation (35%). Now, you can load the spectral data into the NeuroDeveloper software. Perform class assignments and spectral pre-processing. Teach and validate the ANNs and store the network (see Synthon's NeuroDeveloper software manual for details).
- The network file (*.snt) contains all relevant information for pre-processing and classification. This file, and Synthons's run-time environment (NOT the NeuroDeveloper!) are required for classification.

B. Produce NeuroDeveloper maps (e.g. for external validation)

- Load an hyperspectral data file. It is recommended to produce first a map (e.g. chemical map) from original spectra.
- Select '*Synthon maps*' from the '*Image manipulation*' pull down menu. A window entitled '*create NeuroDeveloper maps*' will appear. Press the '*load*' button and select one of NeuroDeveloper's Network files (*.snt).
- Spectra from the data block of original data are now written to a temporary file (in CytoSpec's root folder, please make sure that sufficient free disk space is present). The data are then pre-processed and classified by Synthon's run-time environment. After this, the classification results are automatically transferred back to CytoSpec. When finished, you can immediately press the '*image*' button that causes CytoSpec to display the NeuroDeveloper map. If you wish to modify the NeuroDeveloper exclusion criteria such as WTA, 406040, or extrapolation, uncheck the respective checkboxes and press '*define*'. Change the settings, close the window and press '*image*'. Classification statistics are available by pressing the '*stats*' button of the '*create NeuroDeveloper maps*' window.

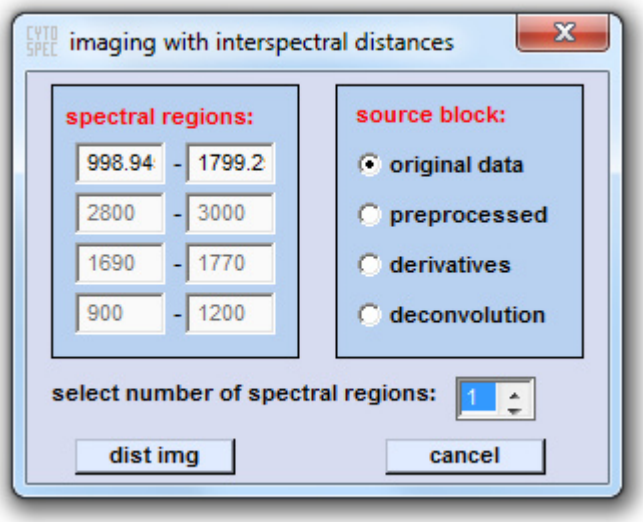
Imaging with Distance Values

Principles of the function: This multivariate imaging function uses multivariate distances between internal or external reference spectra and spectra contained in the active hyperspectral data set. The resulting array of distances has the dimension (xdim, ydim) with xdim being the number of spectra in x-direction and ydim being the number of spectra in y-dimension. Distances are subsequently converted to color scales. A '*distance image*' can be then produced by plotting the colors as a function of the spatial coordinates.

Mathematical distances - related web links:

- Distance (Wikipedia)

When the function '*Imaging with distance values*' was selected from the '*multivariate imaging*' pulldown menu the following window comes up:



The window for data preparation is the same as for other multivariate imaging approaches, such as HCA, PCA, VCA, or KMC imaging. Before this imaging function can be started you have to prepare the data from a spectral multi file.

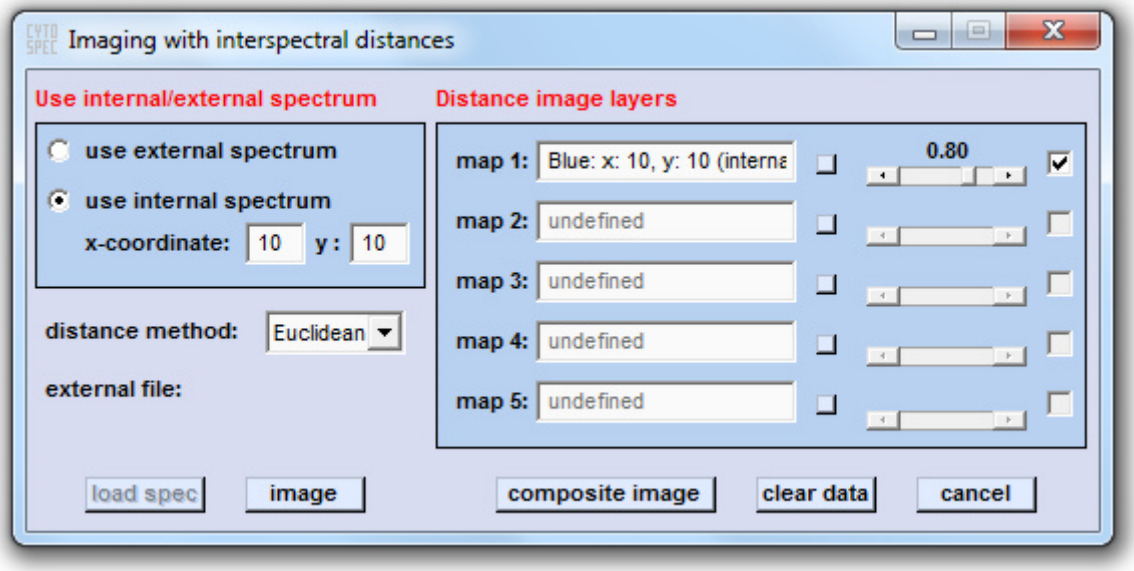
Reformatting data sets for imaging with distance values: The following data manipulations are performed to prepare the hyperspectral data set:

- regions from the selected spectra are selected (data block of your choice),
- spectra are scanned for negative quality test results, or unselected regions of interest, and eliminated.

In the window, select the number of spectral regions (min: 1; max: 4) and choose the data source block by selecting the appropriate radio button (e.g. of the derivative data block). Then, indicate the wavenumber limits of the spectral regions. These regions may NOT overlap, otherwise an error message will be given. To proceed press '*dist img*' or hit '*cancel*' to exit.

Important: Spectra with a negative Quality Test result, or unselected Regions of Interest are automatically excluded from the function '*imaging with distance values*'.

Once data preparation has been finished, the following dialog box will appear:



use external spectrum: when this radiobutton is activated an external double column ASCII spectrum can be loaded. When the file was loaded successfully the directory and the file name are displayed. Press '*image*' to obtain distances between this file and the active imaging data set and to produce a distance image.

use internal spectrum (default): interspectral distances and distance images are produced using an internal spectrum of the current hyperspectral data set. The coordinates of this internal spectrum are displayed by the following two edit fields.

(x,y) coordinates: both edit fields display the (x,y) coordinates of the current internal reference spectrum. The coordinates can be modified manually, or by clicking into one of the images.

distance method: choose between the following interspectral distances: Euclidean, standardized Euclidean and D-values (normalized Pearsons's correlation coefficients).

external file: the file name and path of the external reference spectrum are shown

image distance layers: The function '*imaging with distance values*' allows to produce up to five different distance images. Each of these images can be considered as a layer of a *so-called* composite image that is produced by a superposition of the individual distance images, or maps. The following gui elements are useful to produce the composite image:

> **edit field**: shows the color by which the layer will be displayed in the composite image. Gives also the (x,y) coordinate of the internal spectrum (or the file name of the external reference spectrum). Modification of the content will have no effect on the resulting composite image.
>
> **button**: the composite image is produced
>
> **slider**: changes the contrast of the individual layer in the composite image. The contrast can be varied between 0 (no contrast), 0.8 (default) and 1 (maximum contrast).
>
> **checkbox**: deletes the individual layer from memory.

load spec (button): permits to load the external reference spectrum (double column ASCII).

image (button): interspectral distances between an internal/external reference spectrum an the spectra of the active hyperspectral data set are obtained. Distances are employed to produce a single distance image which is plotted in the axis of the preprocessed maps (main window) using the colormap 'black'.

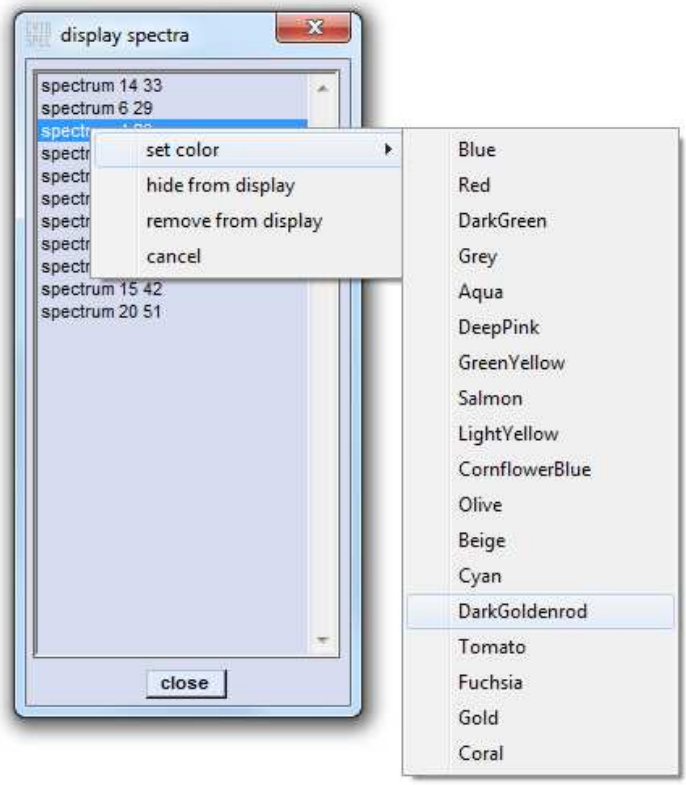
composite image (button): the composite image is produced

clear data (button): all layers are cleared from memory

cancel (button): press this button to exit.

Display Spectra

This option can be used to manage the appearance of individual spectra, or groups of spectra. You can activate the '*display spectra*' window by selecting the option '*display spectra*' from the '*tools*' pull down menu.
Please note: In order to display more than only one spectrum, check the option '*add spectra*' in the Display Option window. Then, you can display spectra of your choice by clicking in the IR maps generated from original or processed spectra. The '*display spectra*' window will show you all active spectra with the x- and y- pixel coordinates. To select/unselect spectra in the '*display spectra*' window use the mouse and hold either key '*Shift*' or '*Ctrl*' pressed (standard windows behavior).



Activate the context menu by selecting one or more spectra and a right mouse click.

set color: allows to set the color of a selected spectrum/selected spectra.

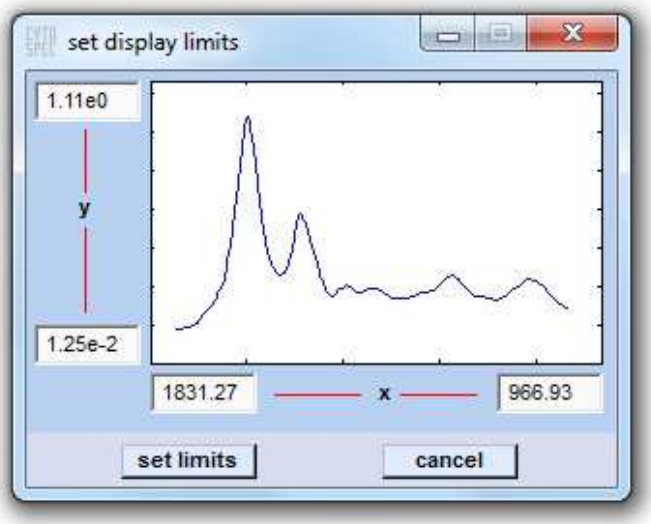
hide from display: hides spectra from the spectral panel. Spectra can be, however, reactivated afterwards.

remove from display: removes spectra permanently from the spectral panel.

cancel: check out what will happen ;-).

Set Display Limits

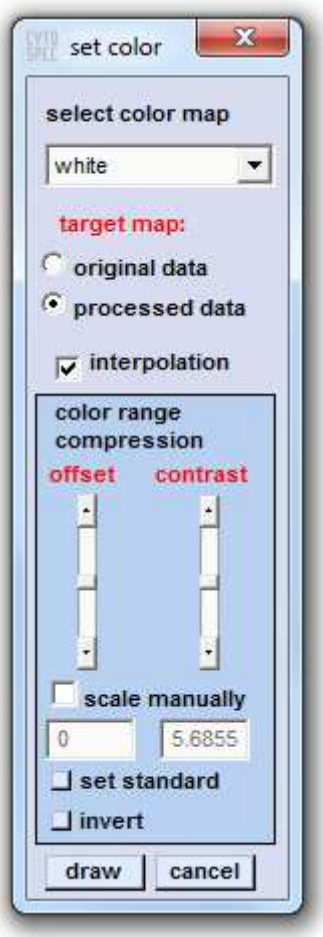
A function that permits to enter display limits of the spectral panel. This window can be activated by choosing option '*set display limits*' from the '*tools*' pull down menu.



Please indicated the new settings for the x- (in wavenumbers) and/or y- (usually in absorbance units) axis. Wrong settings will produce an error message.

Select and Adapt Color Maps

Select and adapt color maps to a given hyperspectral map.



select colormap: allows selection of one of the following color maps: jet, hsv, hot, cool, bone, winter, summer, autumn, copper, prism, flag, lines, colorcube (multicolor mode), ann , blue, red, green, yellow, black, cyan, and magenta (the latter are unicolor modes)

target map: please select whether you want to apply changes to the map reassembled from original or processed data.

interpolation: when this check box is activated, hyperspectral maps are interpolated. Otherwise, the maps are textured (see illustration below).

color range compression: permits you to increase/decrease the offset and the contrast of the color map. The effects of the respective sliders are illustrated in the image below.

scale manually: permits to modify manually the color map by entering the minimal and maximal z-values into the appropriate edit boxes. When this check box is checked the functions '*set standard*', '*invert*', and the sliders '*offset*' and '*contrast*' (color range compression) are deactivated.

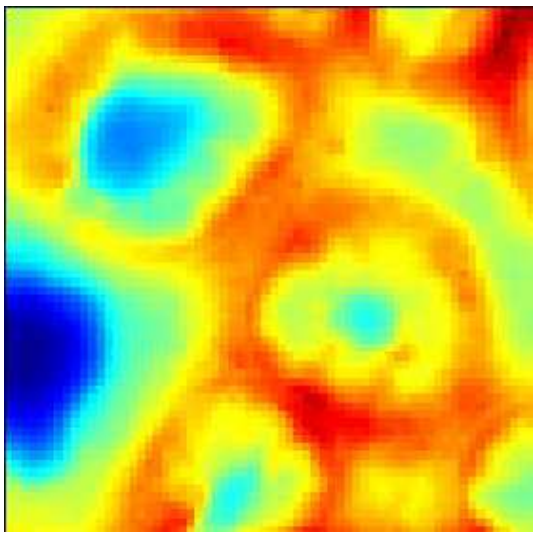
set standard: sets the *offset* and the *contrast* of the colormaps back to default values.

invert: this immediately inverts the colormaps.

draw: applies the changes to the hyperspectral map.

cancel: closes the application

illustration of the effect of the checkbox *'interpolation'*:



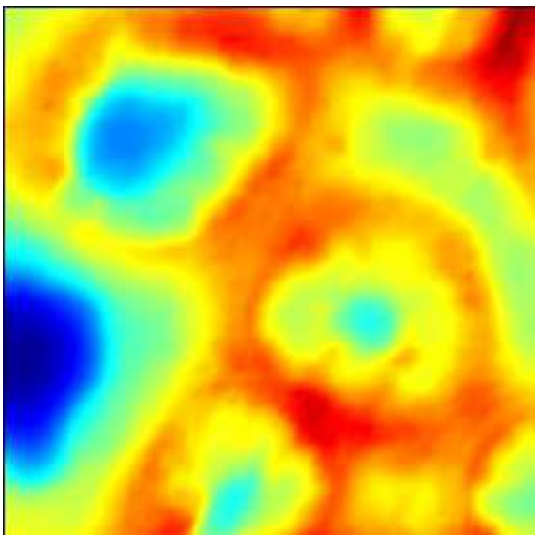
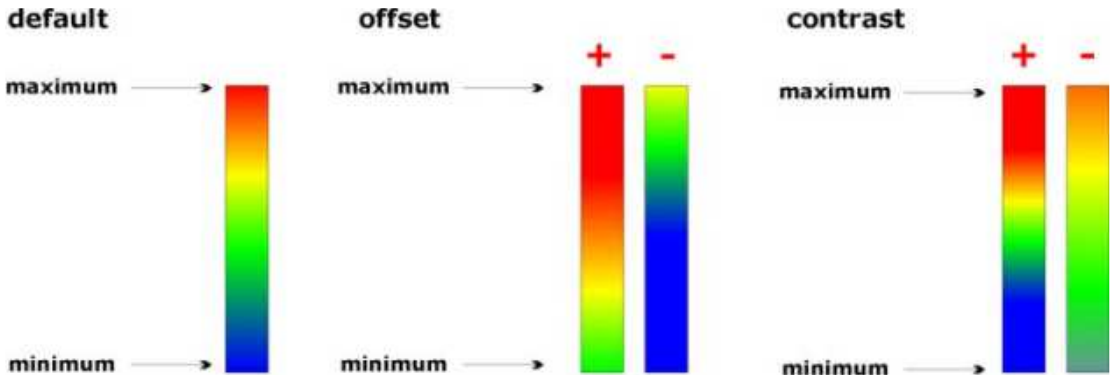


Illustration of the effects of changing the sliders *'offset'* and *'contrast'*:
(example of the colormap jet)



The colormap ann (artificial neural network): This colormap is particularly useful to generate maps on the basis of classification methods producing crisp class membership values, such as:

HCA imaging - images re-assembled on the basis of hierarchical clustering
KMC imaging - images re-assembled on the basis of k-means clustering
Synthon imaging - images re-assembling on the basis of Synthon's NeuroDeveloper(TM) neural network software
ANN imaging - images re-assembled on the basis of the Stuttgart Neural Network Simulator (SNNS)

The colormap **ann** can be individually modified by editing a simple text file '**color.txt**'. This file is located in CytoSpec's root directory, usually in C:\program files\CytoSpec\CytoSpec and is read by CytoSpec upon its initialization. When editing this file, please use only color names given at http://www.w3.org/TR/2003/WD-css3-color-20030214/ chapter 4.3. SVG color keywords. Please check for case sensitivity and spaces!

An example of the content of the file color.txt is given here: color.txt

The settings of the colormap **ann** define also the following color sequences:

1. color options of the '*set color*' function available from the context menu of the Display Spectra function.
2. selectable colors available from the listbox '*color of spectra*' of the Display Options window.
3. the color sequence is furthermore used to display cluster average spectra of the functions KMC, FCM and HCA imaging (button '*spectra*' --> button '*plot averages*').

Grid On/Off

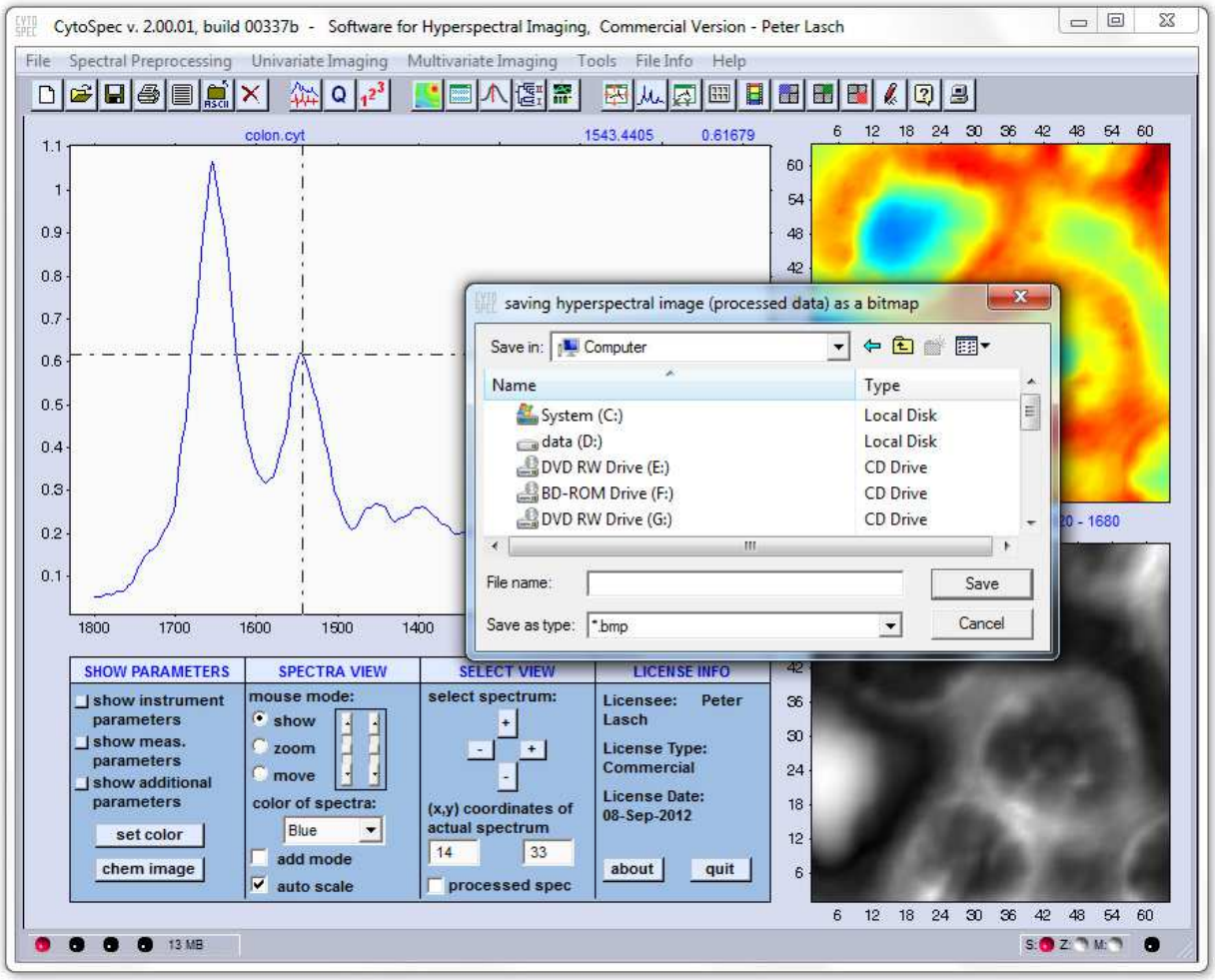
This option is available from the tools pull down menu and has effect on the spectral plot window, only. The grid ON/OFF function may be useful when the spectral window is exported via the Capture --> Spectral Plot function.

Capture

Capturing maps can be started from the *capture* option of the *tools menu*. You can either save the image maps to your hard drive (as bitmaps, map produced on the basis of original, or processed data, spectral window), or alternatively capture the entire program window to the clipboard.

Saving images as bitmaps: Images assembled on the basis of original spectral data (*tools --> capture --> upper image*) and of processed spectra (*tools --> capture --> lower image*) can be captured. When the corresponding option was selected, it is asked to type in a path and a file name. In order to store the spectral plot you have to select the *capture --> spectral plot* option of the tools pull down menu. The option Grid ON/OFF activates or deactivates a grid of the spectral plot ('*tools*' *menu*) The standard resolution used to capture the bitmaps is 300 dots per inch (dpi). Files are stored in a standard windows bitmap (.bmp) data format.

Capturing program window: Along with storing images or spectra you may also want to take a screenshot of the program window. By choosing the *capture --> window* option of the *tools* pull down menu and selecting the resolution (72, 150, 200, or 300 dpi) the entire CytoSpec program window is copied to the clipboard.



Export Maps

This function can be used to export the z-data of hyperspectral maps. The z-data are stored in standard ASCII text files. To export the data you have to select first the option '*export maps --> upper plot (or lower plot)*' from the '*tools*' pull down menu. Then, in a standard windows dialog box you will be asked for path and file name of the data file to be stored.
Data are stored in files of the following structure (example of a map of M x N pixel spectra, [xdim x ydim]):

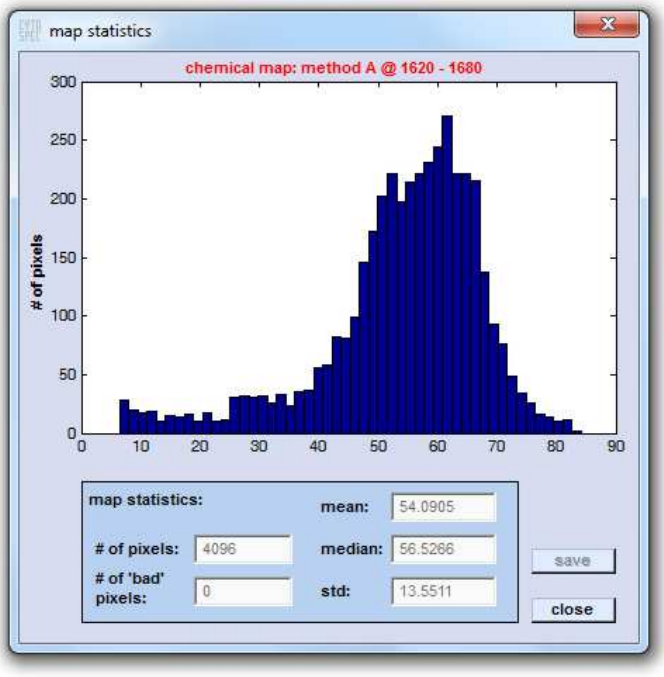
	1	**2**	**3**	**4**	**.**	**N**
1	0.2	0.1	0.2	0.1	.	0.1
2	0.4	0.7	0.4	0.7	.	0.7
3	0.5	0.4	0.5	0.4	.	0.4
4	0.7	0.2	0.7	0.2	.	0.2
5	0.8	0.4	0.8	0.4	.	0.4
6	0.9	0.9	0.9	0.9	.	0.9
7	0.7	0.7	0.7	0.7	.	0.7
8	0.4	0.4	0.4	0.4	.	0.4
9	0.2	0.8	0.2	0.8	.	0.8
10	0.6	0.2	0.6	0.2	.	0.2
.	.	.	.	.	.	.
M	0.6	0.5	0.3	0.1	.	0.9

The ASCII-files are now available for import into other programs such as Excel, Origin or Matlab. Note that the standard extension of the ASCII data files will be *.dat.

Map Statistics

This function can be used to visualize the distribution of the actual z-data (e.g. intensities) as a histogram. The option '*map statistics*' is available from '*tools*' pull down menu. You can do map statistics from maps assembled from original data (upper hyperspectral image), or from maps from any type of processed data (lower map).

# of pixels: displays the total number of pixels of the spectral map.
# of bad pixels: displays the number of pixel spectra that had failed one of the quality tests.
mean: displays the mean value of the data used for imaging.
median: shows the median of the image data.
std: the standard deviation .



Display Large Maps

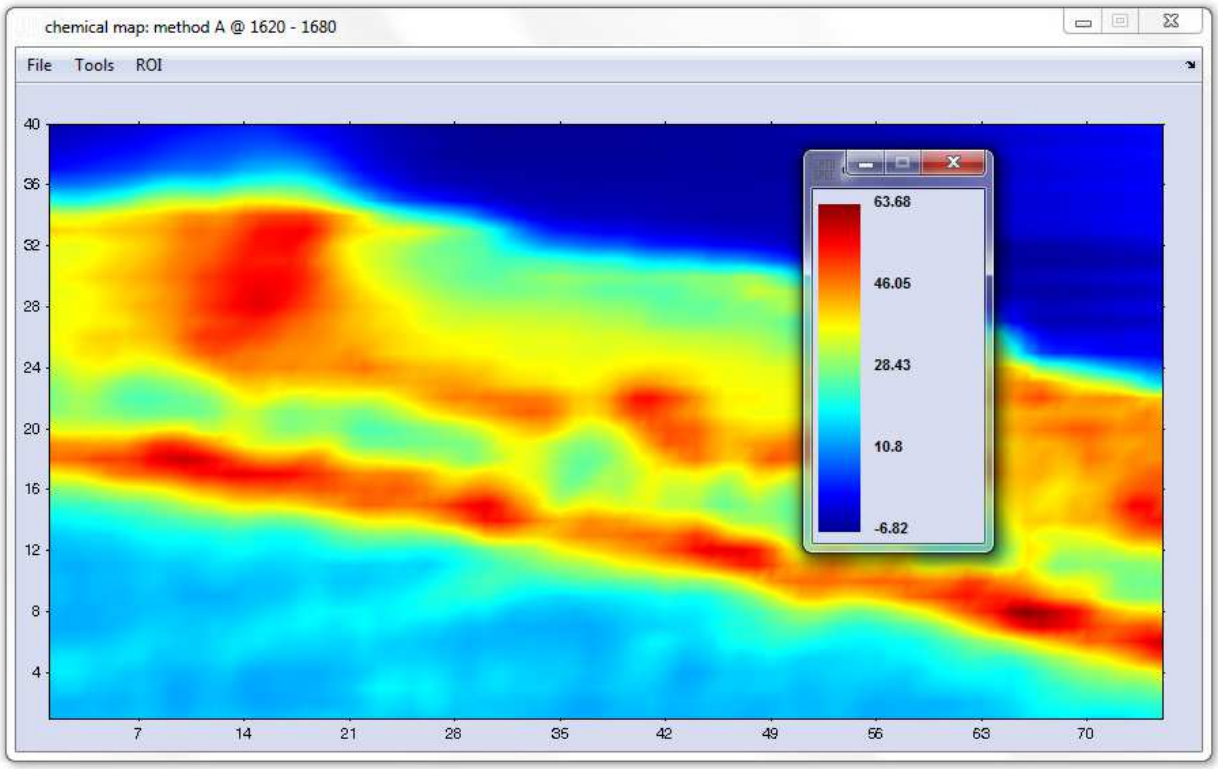
The function '**display large maps**' is available from the '*tools*' pull down menu but also from the context menus of hyperspectral maps (click by the right mouse button). When this option was chosen, a new window will appear showing the maps in the true spatial aspect ratio.

The following information can be obtained from '*large maps*':

1. the pixel position of the cursor,
2. (x,y) - spatial position of the cursor (obtained from the parameters SZX and SZY)
3. the z-value (absorbance, transmittance, Raman intensity etc.) found at the current cursor location.

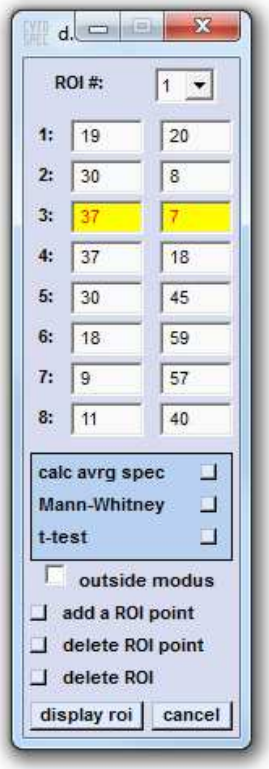
Further options:

1. the map can be stored as a bitmap by selecting '*capture to bitmap*' ('*tools*' pull down menu).
2. a colormap is available by selecting '*display colorbar*' ('*tools*' pull down menu).
3. the window can be deleted (choose '*exit*' from the '*file*' pull down menu).
4. one can define Regions of Interest (ROI)



Define Region of Interest (ROI)

The function **'define region of interest** (ROI)' is available from the '*tools*' pull down menu but also from context menus of hyperspectral maps (click right mouse button). When this option was chosen, a new window Large Map will appear that shows the map in the real spatial aspect ratio. Furthermore, an additional window '*define ROI*' pops up which permits to define up to three different regions of interest in one hyperspectral map.



listbox **ROI #**: Up to three different regions of interest can be defined. Switching between active ROIs can be achieved by selecting the appropriate option from this listbox.

edit fields **1: - 15**: Please enter the (x,y) coordinates of ROI boundary points (in pixel coordinates) in the respective fields. Alternatively, you can click into the 'large map' - the pointer coordinates will be transferred automatically.

calculate average spectrum: a tool that permits to calculate/display and store average spectra from all active regions of interest. Works in exactly the same way as the option '*spectra*' described in the HCA imaging, FCM imaging and KMC imaging functions.

Mann-Whitney test: not yet implemented

t-test: not yet implemented

checkbox **outside modus**: this option permits to switch between the in- and outside modus of the ROI definition procedure.

checkbox **add ROI point**: fields that permit to define additional ROI boundary points are added.

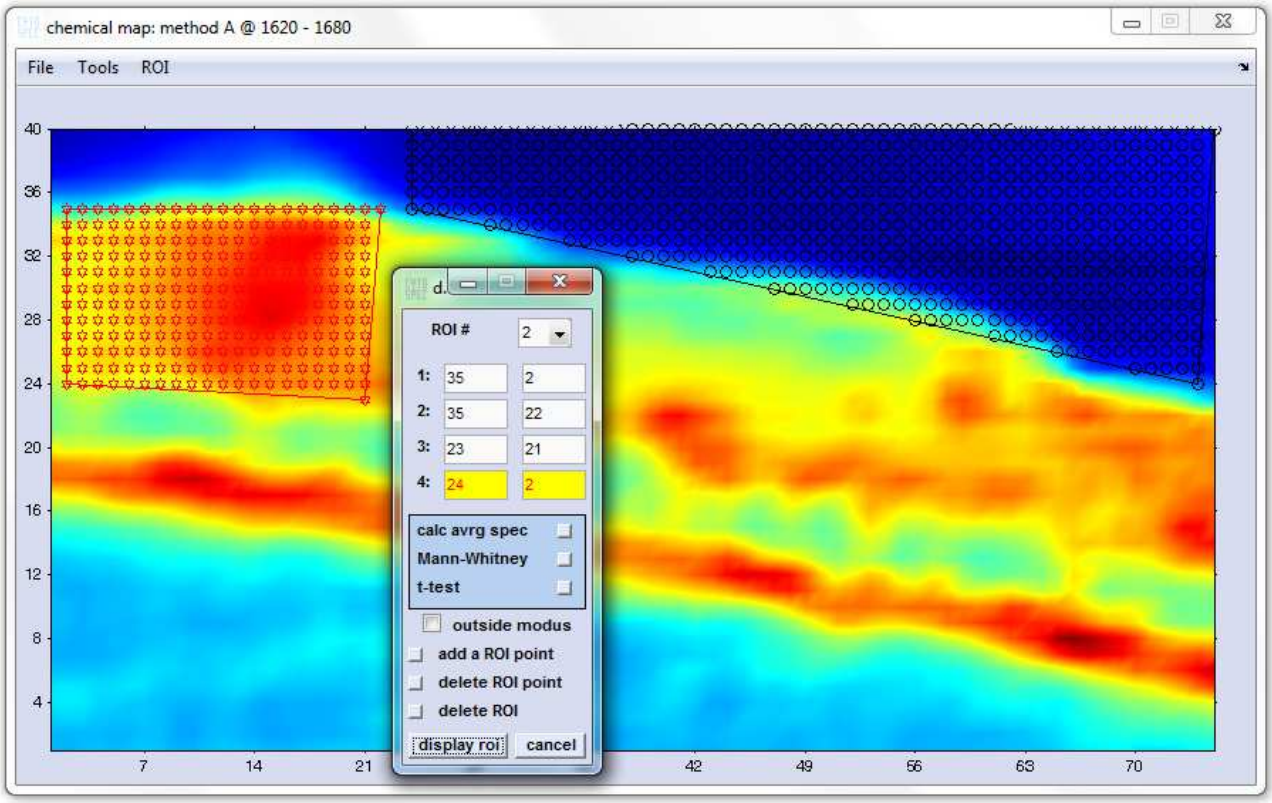
checkbox **delete ROI point**: ROI points are deleted.

checkbox **delete ROI**: all ROI point definitions made so far are deleted.

display ROI causes to draw the region of interest in the window '*large map*'

cancel the routine is aborted

When a region of interest has been marked within the hyperspectral map the option '*apply ROI --> preproc data*' from the pull down menu of the large window becomes activated (see screenshot below). The function '*apply ROI --> preproc data*' creates a new set of preprocessed spectra from the data block of original spectra by transferring only spectra located within the actual ROI area. Spectra from areas outside are omitted and filled by NaN-values (Not a Number). Consequently, spectra originating from non-ROI areas are excluded from all further image processing and appear in any hyperspectral map generated on the basis of preprocessed data as black pixels. Thus, the '*apply ROI --> preproc data*' function works similar to a manual Quality Test .

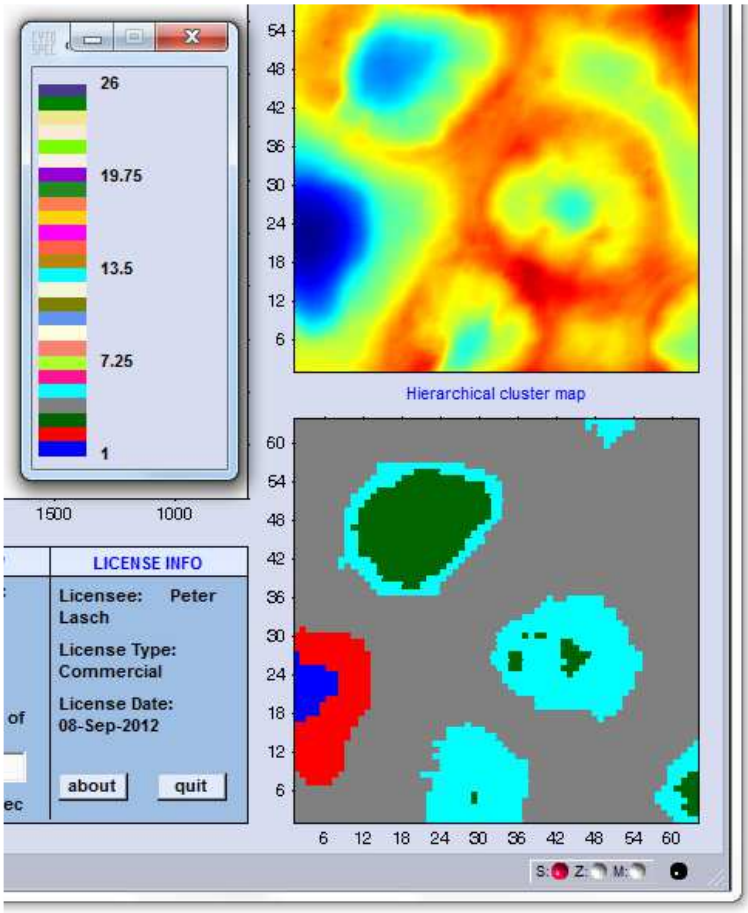


Display Colorbar

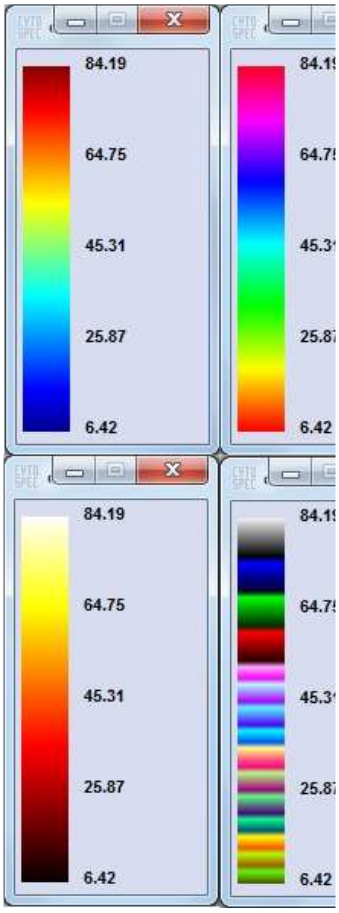
This function is available from the '*tools*' pull down menu and displays a colorbar of the active map. It use the active colormap which is available from the Set Color function of the '*tools*' pull down menu. Colorbars can be obtained for all types of maps, including Large Maps) and maps produced on the basis of crisp class membership values such as

HCA maps - images re-assembled on the basis of hierarchical clustering
KMC maps - images re-assembled on the basis of k-means clustering
Synthon maps - images re-assembling on the basis of Synthon's NeuroDeveloper(TM) neural network software
ANN maps - images re-assembled on the basis of the Stuttgart Neural Network Simulator (SNNS)

Example of a colorbar for a KMC map. Note that in the example below the active colormap is of type **ann**. Although only five clusters are displayed, the colorbar will always show the complete colormap which consists here of 27 different colors. Note furthermore, that the sequence of colors is defined by the numbering of colors in the text file 'color.txt' (can be found in the root directory of CytoSpec, usually C:\program files\CytoSpec\CytoSpec). For details, see ann Colormap

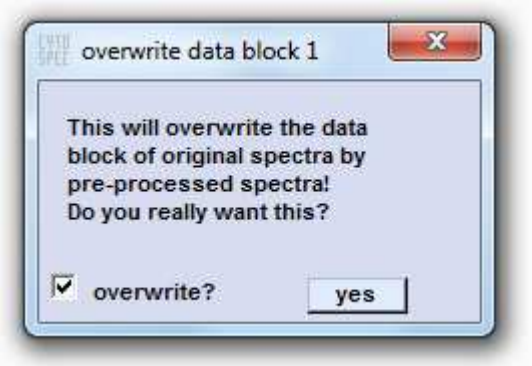


Four examples of the colorbars of the types jet (top left), hsv, colorcube and hot (clockwise direction).



Swap Data Blocks

Overwrites the original data block by preprocessed data. This function is rarely required; however, the test for thickness of the Quality Test (which can be performed exclusively on original data) may require offset corrected spectra. As original data are lost it is highly recommended to store the data before performing the '*swap data block*' function. If the block of preprocessed data is empty CytoSpec returns an error message. To avoid overwriting of the original data as a result of typing error, CytoSpec asks for confirmation (see window below).

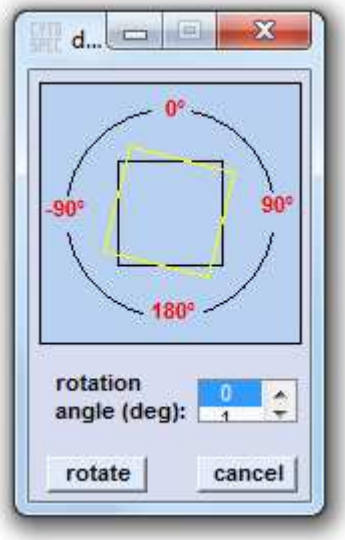


Rotate & Flip

Data blocks AND images are manipulated such that the spatial dimensions are changed (i.e. swapping the x- and y-dimensions) while the spectral dimension remains unaffected.

Rotate function can be used to rotate the data arrays by 90 degrees clockwise, counterclockwise, or by 180 degrees, respectively.

The *Flip* function permits to flip the spectral hypercube along the x-axis ('*flip horizontally*'), or alternatively along the y-axis ('*flip vertically*').



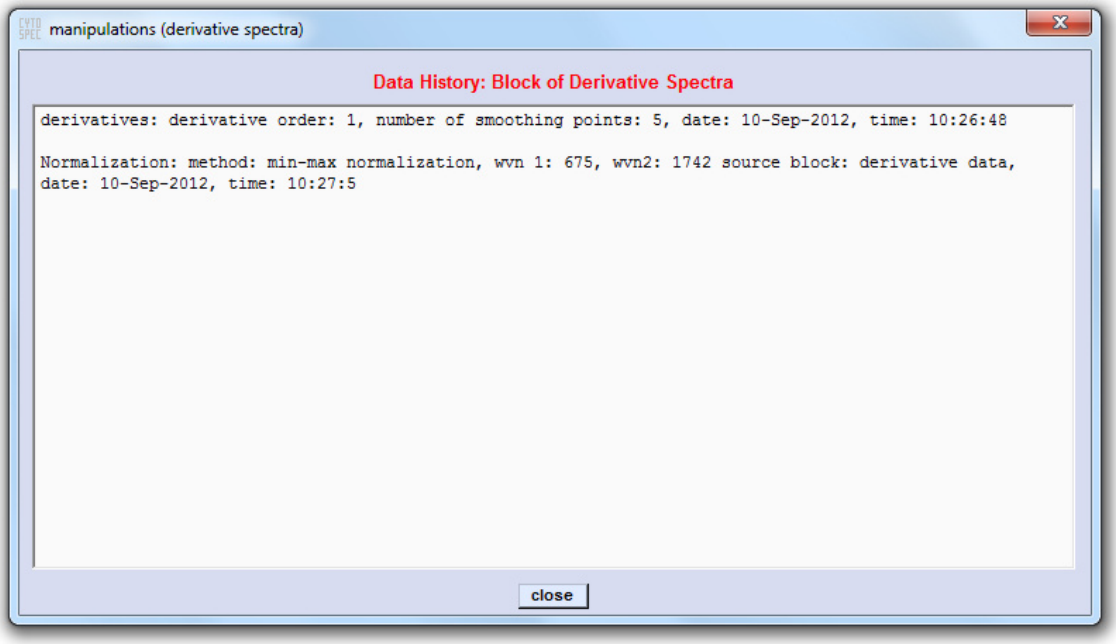
Show Data Block History

Show data block history: The history of data manipulations is stored within the spectral multifile. In order to view this information select '*File info* → *Show history*' and select the data block of interest.

The information includes:

1. data source, time and data of import (block of original data).
2. time and date of data block creation (data block of preprocessed, derivative, and 3D -FSD spectra)
3. parameters of preprocessing (derivatives, cut, smooth, interpolate, baseline correction, quality test, etc.)

Additional information is available from the log-file.



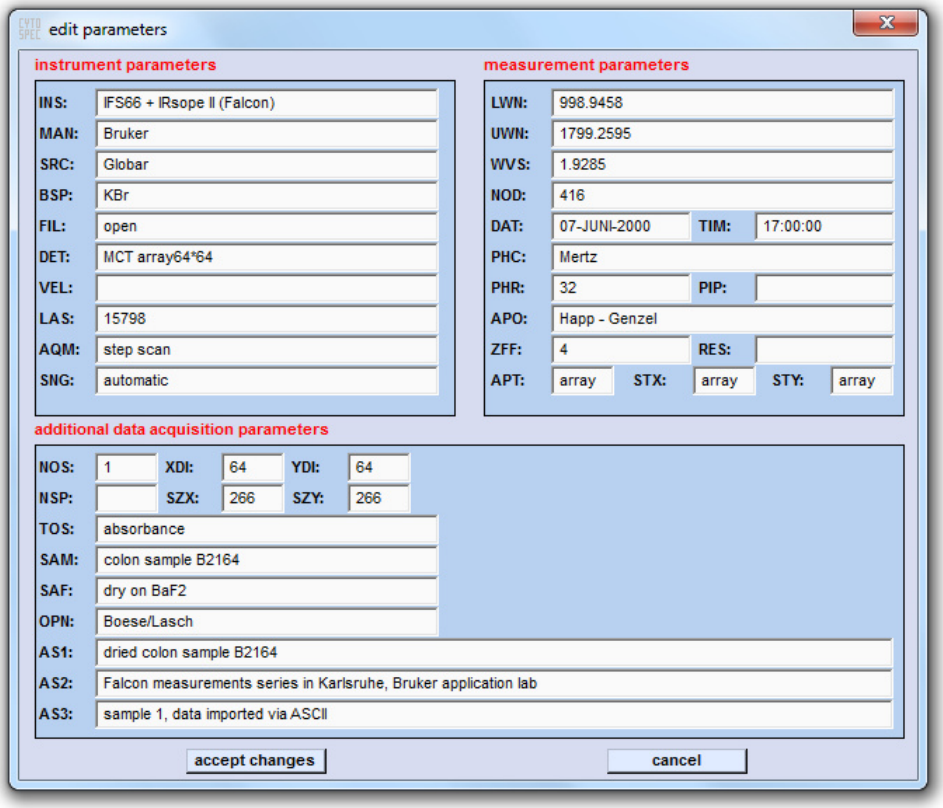
Show Parameters

Displays parameter of the following data parameter blocks:

- Instrument parameters
- Measurement parameters
- Additional data acquisition parameters

Detailed information on the data parameters is available in the following chapter of this help manual. Additionally, the data parameters can be modified using the function Edit Parameters of the '*File Info*' pull down menu.

Edit Parameters



Instrument parameters:

INS - instrument type
MAN - instrument manufacturer
SRC - source setting
BSP - beamsplitter setting
FIL - type of optical filter
DET - type of detector
VEL - scanner velocity
LAS - laser wavenumber
AQM - acquisition mode
SNG - signal gain

Additional data acquisition parameters:

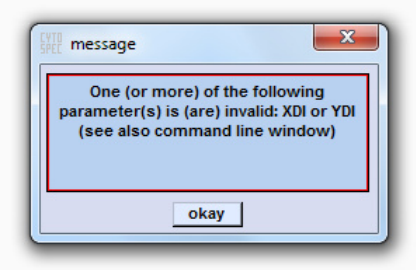
TOS - type of spectra
NOS - number of scans
XDI - number of pixel in x-direction
YDI - number of pixel in y-direction
NSP - number of spectra
SZX - area size in x dimension (µm)
SZY - area size in y dimension (µm)
SAM - sample name
SAF - sample form
OPN - operator name
AS1 - additional sample information 1
AS2 - additional sample information 2
AS3 - additional sample information 3

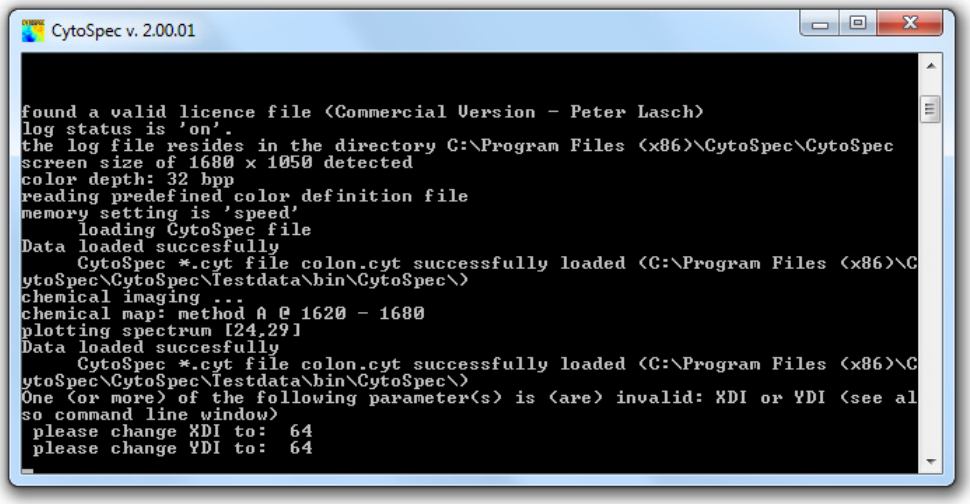
Measurement parameters:

LWN - starting wavenumber
UWN - ending wavenumber
WVS - wavenumber step
NOD - number of spectral data points
DAT - date of measurement
TIM - time of measurement
PHC - phase correction mode
PHR - phase resolution
PIP - phase interferogram points
ZFF - zero filling factor
RES - nominal resolution
APO - apodization function
APT - aperture size (µm)
STX - step in x dimension (µm)
STY - step in y dimension (µm)

Parameters highlighted by **red color** should be carefully modified since these parameters are used by the software. Modification of these parameters may cause malfunctioning.

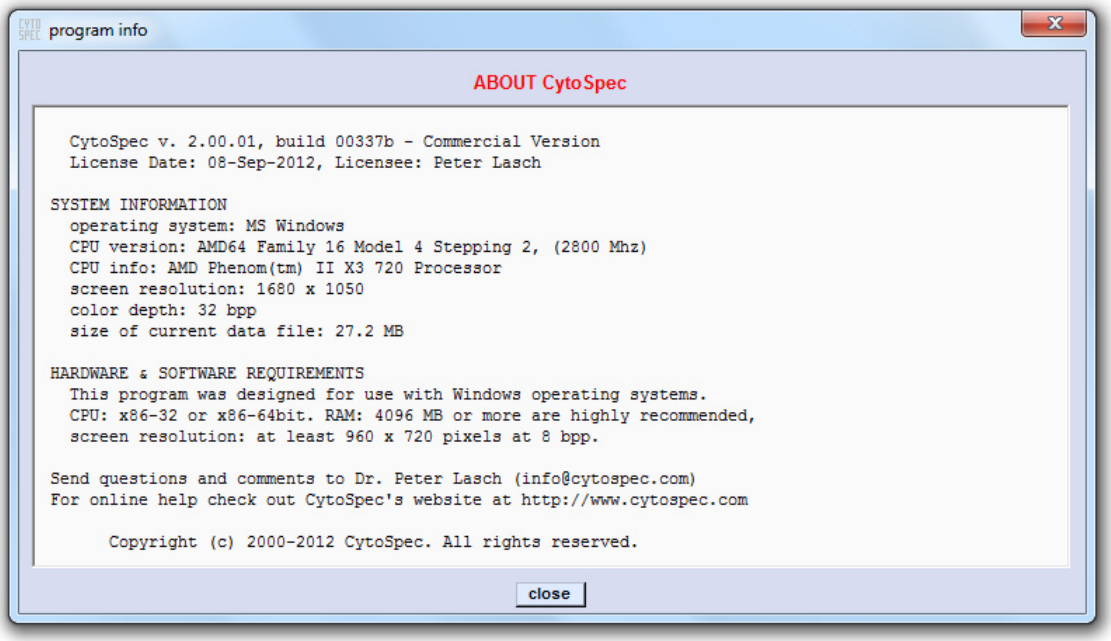
Changes of the parameters highlighted by red color are tracked and checked when leaving the '*edit parameter*' window. In case of errors or data inconsistencies an error messages will be displayed. The CytoSpec program also makes suggestions of how to change the faulty parameters (see command window below).





About

This function displays the CytoSpec program version number and basic license information.



Help - Description

This function displays the CytoSpec online help. To obtain the most recent version of the CytoSpec online help go to the CytoSpec web pages at

http://www.cytospec.com