COMSTAT

USER MANUAL

(REVISED 14/8-2010)

Written by
Arne Heydorn
aheydorn1@gmail.com

COMSTAT is a program for quantification of three-dimensional biofilm structures. It analyzes stacks of images acquired with confocal laser scanning microscopy (SCLM). It was written as a script in MATLAB 5.3 (The MathWorks Inc., Natick, Massachusetts), equipped with the Image Processing Toolbox. The program is menu controlled, user-friendly, and requires no prior knowledge in programming or image analysis. MATLAB 5.3 equipped with the Image Processing Toolbox must be available on your computer. Later versions of MATLAB will also work fine. Many universities have special campus licences for MATLAB. You can check out the MathWorks homepage at www.mathworks.com for details on how to obtain MATLAB.

1. Installing and using the programs

The COMSTAT package (i.e. all the files you receive) contains five programs (COMSTAT, CHECKALL, LOOK, LOOKTIF and CONVERT000) and a number of functions used by the four programs. Place all the files in the folder MATLABR11/work. If this folder exists on your computer you don't have to do anything else. Alternatively, place the program files in any folder you desire and add this folder to the path in MATLAB (see the MATLAB help desk on how to add a folder to the MATLAB path). Each of the programs is started by typing the name of the program on the command line in MATLAB and pressing *<ENTER>*. However, the programs must be started in the folder containing the image stacks. Use the MATLAB commands *dir, pwd, cd .., cd <folder>* to change to the directory containing the images. Any of the programs can be stopped during execution by pressing *<CTRL><C>*.

2. Image formats

COMSTAT analyzes image stacks acquired with SCLM. Unfortunately, different confocal microscopes (Leica, Zeiss, Olympus, Nikon, Biorad) usually store image data in different formats. COMSTAT was originally designed to analyze images from the Leica TCS4D confocal microscope. However, COMSTAT can also analyze images in other formats, which will be described below.

At present, COMSTAT requires all images of a stack to be stored as individual '.tif' images. The '.tif' images can be of any sizes (e.g. 256×256, 512×512 and 1024×1024), and be 8-bit, 16-bit or 32 bit. However, all of the programs will convert the images into grey-scale 8-bit images. Using 8-bit 512×512 images is fine for the purpose of quantifying biofilm structures. COMSTAT requires an '.info' file for each stack of images. An '.info' file is a text file containing information about the image stack. COMSTAT uses informations from the '.info' file when analysing the image stack, and it is therefore essential that the '.info' file is present. The Leica TCS4D microscope automatically generates an '.info' file when an image stack is saved. The Leica TCS4D '.info' file contains image information such as the time and date of acquisition, pinhole size, pixel sizes, etc. COMSTAT reads informations about the image stack from the '.info' file. However, if you do not have a Leica TCS4D microscope you can still write your own '.info' files, which can be read by COMSTAT. By using a text editor (e.g. Notepad) you can write a text file with the extension '.info'. This '.info' file must look like the following example:

```
Range
#11#
#13#
Pixelsize(x)
```

#0.49#
Pixelsize(y)
#0.49#
Pixelsize(z)
#1.52#

The range tells COMSTAT how many images this stack contains. The pixelsizes in the x, y and z directions are also necessary. The pixelsize is the distance between two neighbouring pixels and is given in micrometer. The most important thing about this type of '.info' file is that it contains the data between the # signs. COMSTAT locates the data by looking for the # signs. Moreover, the images belonging to a certain '.info' file must be named accordingly. If the '.info' file above was called hello.info, the images belonging to this '.info' file should be named:

hello11.tif
hello12.tif
hello13.tif

The range can start from any positive integer value (e.g. 1 to 10 or 232 to 245). In case there is only one image in a stack, the pixelsize(z) should be set to zero.

Image stacks named xxx001, xxx002 xxx003 etc.

Some confocal microscopes generate image stacks numbered in the following way: xxx001, xxx002, xxx003, xxx004, xxx005, xxx006, xxx007, xxx008, xxx009, xxx010, xxx011, etc. This format is not recognized by COMSTAT. However, the program CONVERT000 can change the names of the images to xxx1, xxx2, xxx3, xxx4, xxx5, xxx6, xxx7, xxx8, xxx9, xxx10, xxx11, etc. which is recognized by COMSTAT. In order to rename your images you first have to write an '.info' file named xxx.info as described above. Then you start the CONVERT000 program in the folder containing the images and the images will be renamed accordingly.

Other image formats than those described above

If the images from your CSLM are named differently than those described above, you have to manually rename your images to a format compatible with COMSTAT. Since this can be very time consuming you might want to write a small MATLAB script similar to the CONVERT000 program in order to change the names of your images automatically. If this is a problem, you are welcome to contact the author, and I may be able to help you

3. Checking the image stacks by the CHECKALL program

The CHECKALL program checks if all the image stacks are intact. CHECKALL checks each individual image stack for missing images or damaged images. It also checks for errors in any of the '.info' files. It is a very good idea to run the CHECKALL program and correct any errors before doing any further analyses.

4. Determination of threshold values by LOOK and LOOKTIF

In order for COMSTAT to analyse the image stacks, the threshold values must first be determined. The need to do thresholding is due to the fact that the images coming from the microscope are grey-scale images, and COMSTAT does not automatically know which pixel

values should be taken as biomass and which pixel values should be taken as background. Thresholding an image stack in COMSTAT results in a three-dimensional matrix with a value of ONE in positions where the pixel values in the original image are above or equal to the threshold value, and ZERO where the pixel values are below the threshold value. The value ONE represents positions containing biomass, while ZERO represents the background. The program LOOK can be used to determine the threshold value of each image stack. By doing simple thresholding with different values, the resulting images can be scrutinized and a suitable threshold value can be determined. Often the same threshold values can be used for images taken by the same person on the same day, whereas the threshold values frequently change from day to day or when another user is doing the image acquisition. When the program LOOK is used, it can sometimes be difficult to see the small details in the images because a large number of images are present on the screen at the same time. In that case, the program LOOKTIF may be used instead. LOOKTIF imports individual tif images and can do simple thresholding like the LOOK program. Write down the threshold values for the different image stacks, so they can by typed into COMSTAT later.

5. Using the COMSTAT program

If you have run the CHECKALL program on your image stacks and they were okay and you have determined the threshold values by the LOOK and LOOKTIF programs, you are now ready to analyze the image stacks by COMSTAT. The COMSTAT program has been described in two papers (Heydorn *et al.*, *Microbiology* 146(10):2395-2407; Heydorn *et al.*, *Microbiology* 146(10):2409-2415). Most of the variables that can be calculated by COMSTAT are described in these papers.

When COMSTAT is started it first lists all image stacks found in the present directory and the user is asked which of the image stacks should be analyzed. Then the threshold values (which were determined by using the LOOK and LOOKTIF programs) must be entered. When the threshold values have been entered, a menu appears containing all the choices available in COMSTAT. The first section of options (1-11) is the range of image analysis tools that can be used in analysing the image stacks. These image analysis tools are described in the two papers mentioned above. Note that the image analysis functions 1-5, 8, 10 and 11 are performed very fast whereas the functions 6, 7 and 9 are quite slow. Therefore, options 6, 7 and 9 should not be chosen if these variables are not really needed.

Option 21. By choosing this option, the preprocessed images, i.e. the thresholded and connected volume filtered images are saved in a separate folder on the computer. In this way, the pre-processed images can be checked after the analyses.

Option 22. By choosing this option, images that are used and calculated during the analyses will be displayed on the screen during program execution. This makes the COMSTAT program slightly slower.

Option 23. By choosing this option, the computer will correct for a non-flat first image. It is very important for many of the calculations in COMSTAT that the first image of the stack shows exactly the layer of biomass at the substratum. However, often the coverslip is not entirely plane or the coverslip is tilted. This means that the first image only contains some of the biomass at the substratum while the rest is in the second or even the third image of the stack. This problem can easily be seen with the LOOK or LOOKTIF programs. COMSTAT provides a solution to this problem. If option 23 is chosen, COMSTAT does the following. If the biomass coverage in the first image is smaller than in the second image, the first image will be replaced with a maximum image of the first couple of images. The user determines the number of images that are used to generate the maximum image. The result is that each pixel value in the first image is the maximum of the pixel values in the same

positions of the first couple of images in the stack. Note that COMSTAT will only change the first image if it suspects that there is a problem, i.e. if the biomass coverage in the first image is smaller than in the second image. Therefore choosing this option does not make any changes to the images, which do not have a non-flat image.

Option 24. Choosing this option, turns off the connected volume filtration of the image stacks. By default, COMSTAT performs connected volume filtration on all image stacks, unless option 24 is chosen. Connected volume filtration is a smart way of removing noise from the images. It works by removing biomass that is not in some way connected to the substratum. This seems reasonable when working with continuous-flow flowcells, where biomass that is not connected to the substratum would be expected to be washed away. However, there may be conditions where connected volume filtration is not desired. If option 21 is selected, the thresholded and connected volume filtered images can be viewed after the analysis. For more details on connected volume filtration, see (Heydorn *et al*., *Microbiology* 146(10):2395-2407; Heydorn *et al*., *Microbiology* 146(10):2409-2415).

Option 25. During execution, the COMSTAT program generates text files containing the results from the quantifications. The results are continuously written to the text files. The text files can be opened in any text reader or Excel for further processing of the results. The numbers in the textfiles are by default point separated. If your Excel program recognizes comma as a decimal symbol instead of point, the text files should be converted to comma format. By choosing this option, the text files are converted to comma format.

6. Getting the results from COMSTAT

The results generated by COMSTAT are continuously written to text files during program execution. This means that if an error occurs which makes COMSTAT stop before having analyzed all image stacks, the text files will still contain the results calculated until that point. An important point is that if you have asked COMSTAT to save the text files in a folder, which already contains text files from another quantification, the new results will simply be appended to the end of the already existing text files. This is an advantage if the images are from the same session or biofilm, because only a single set of text files will be generated. For each of the 11 image analysis features selected, a text file containing the result of this particular feature will be generated. For some of the functions, more than one text file will be generated. Furthermore, a text file *report.txt* is generated containing information about all the images and results from all the selected image analysis features.

If you have questions or comments, check out the COMSTAT homepage at www.imageanalysis.dk or send an e-mail to the address below. Information about errors and suggestions for improvements are very welcome.

Arne Heydorn
Email: aheydorn1@gmail.com