Carmen Garcia-Comas
Written for Zooprocess version 6.16
February 2010, Hydroptic & LOV

SHORT MANUAL TO WORK WITH THE ZOOSCAN

Please use the following "tools" to learn how to use the ZooScan:

ZooScan Manual (v.6.16) to download in www.zooscan.com
ZooScan Users Forum: http://zooscan.forumakers.com/
Gorsky et al. (2010) ZooScan methodological paper.
Open access link in www.zooscan.com
http://plankt.oxfordjournals.org/cgi/reprint/fbp124?ijkey=gJnvV9AjR20gITg&keytype=ref

ZOOSCAN WORKING PROTOCOL

Morning:

Turn on the ZooScan and rinse the scan tray.

Eliminate marks on the glass and frame and check from time to time if the glass of the ZooScan cover has not marks.

Put some water to cover the tray (it avoids scratching the tray with the frame).
Place the frame of your project.

Fill with water until the step of the frame is covered.

Do a background scan (2scans) in the project you will work on (check that there is not dust on the OD, dark circle, position and that the tray and water are clean).
DON'T FORGET TO PRESS THE TEMP GREEN LIGHT BUTTON before scanning.

Day:

Scan samples (while one sample is being scanned you can recondition the previous one or prepare the next one if you want to).

COVER THE STEP OF THE FRAME WITH EXTRA WATER!

PRESS THE TEMP GREEN LIGHT BUTTON before scanning!
(if your ZooScan has also a blue light button and it's on, turn it off before scanning)

End of the day:

Launch “Convert and process images in batch mode” in ImageJ, Z, (All images in the Raw folder will be processed during the night). If in your project you already have processed images, you can select to process the converted images (in the subfolders of the Work folder) to spare conversion time.

Once a week:

You can dedicate 1/5 days to creating your Learning set at the beginning (at the end of the first week of scanning) (we recommend not investing too much time in building it) and to validate your samples (check automatic recognition by extracting vignettes according to prediction and resorting them if wrongly recognized).

We recommend to use a small Learning set (about 200 objects/group) to accelerate organisms recognition, and to validate all the samples to have a robust dataset at the end.

ZOOSCAN PROTOCOL FOR SCANNING

DO NOT FORGET TO DO A BACKGROUND (2 scans) every morning before a day of scanning.

If it is the first time that you will use your ZooScan:

1. Check that your ZooScan is on a bench without vibrations and that the instrument is horizontal.

Check first that the top-left corner of the ZooScan is stable (push on it). If not, you should lift that leg a bit until it does not sink under the pressure of your hand. If the ZooScan is not horizontal, level it by screwing or unscrewing the legs.

2. The height of your ZooScan legs must be enough to place below the mouth of the zooscan a receptacle to recover your sample when pouring it from the scanning tray.

3. The receptacle to recover your sample should be wide enough to not miss the mouth of the tray when elevated, and deep enough (although lower than the ZooScan height) to recover all the water of the sample+ the extra-water poured to rinse the tray.

First time working with ZooProcess:
Read section 12.9 in page 90-91 of Zooprocess_manual_v6.16

If you have just installed Zooprocess for the first time or also if you installed it but it is the very first time that you create a project for your new ZooScan, you need to change the default sampling settings. Check the version of your ZooScan and dimensions of your frames in the Qualification Report that Hydroptic delivered to you with your instrument.

Choose the option "CHANGE/CHECK Zooscan version & parameters" in Zooprocess.

- Select your ZooScan version to the one corresponding to your model
(if delivered by Hydroptic check if it is v1 or v2 in the qualification report).
- Check if the OD position and frame X-Y dimensions correspond to those noted in the Qualification report and if not change them in the computer according to the values delivered by the company.

Now the standard .ini files are modified according to your machine, and your ZooScan is ready to scan.

Create a Project:
Read section 12.9 in pages 90-91 of Zooprocess_manual_v6.16

Open Image J, click on Z icon and choose the option "create a new project" which is on the bottom of the options' list. Click OK.
Choose the drive. We recommend to create your projects in a drive root were you just can store them and you just keep the ZooScan projects (clean) (i.e., Zooscan root folder in C:, and zooscan projects in D: drive).

We recommend that each project has only one scanning configuration file (vuescan.ini file). So choose just the one you will always be using in that project.
It is recommended to use the large frame and highest resolution image: "vuescan_zooscan_2400dpi_frame2_large.ini" (although you need at least 2G RAM to be able to process those images!).
Untick the other options (other .ini files). Only the ticked .ini file will be copied to your Zooscan_config folder in the created project.

Before scanning:

We recommend keeping bottles with water to have litres of water at environmental temperature to work with your zooscan. If not, you will probably have condensation, bubbles, etc, due to temperature differences between the tap pipes and the zooscan room.

Treat the ZooScan gently and maintain it (e.g., clean it at the end of the day; hold always the tray when you lift it to remove the water! It could fall). Use only wooden sticks to separate organisms in the tray (it avoids scratches). We use dried cactus spines stuck with scotch taper on a thin pipette in the Laboratory of Villefranche.

Check the performance of the ZooScan by doing a background scan (2 replicates) every morning. This is also important for image processing, as the process will be done with the latest background (if it was done a long time ago it could maybe not represent the present background of your ZooScan).

4. Turn on your ZooScan if it's the first sample in the morning (it needs to warm up). Then you leave it on all the day if you are passing samples. Turn the zooscan off during the night.

5. Prepare your sample. Here, I explain the procedure followed in Villefranche (for 200 and 330 µm sampling nets):

    1) Sieve your sample to take out the formol and sea water (you can keep it to recondition the sample afterwards). We sieve each sample through a 1mm mesh and 200µm mesh to have 2 size fractions (this is to avoid missing the rare large organisms while splitting the sample to scan a subsample). Rinse the sample with tap water to properly eliminate the sea water. Place each fraction (subsample) in tap water.
    2) Now you have 2 size fractions. The large one (>1mm) is called d1; the small one is called d2 (200µm-1mm) (d1 and d2 will be added at the end of the sample name to distinguish them during the data processing).
    3) Take one of the fractions (e.g., d2) and split it until there are ~1000-1500 individuals (experience will give you the good eye for dilution to have a good scan: not too crowded and not poor).

Scanning (we scan during the day, and process them later during the night):

1. Pour some water on the scanning tray until you have covered it.

2. Place the frame (the one defined in your Zooscan Project; preferably the large one). Control that the frame is well placed on the foremost left-bottom side of

the scanning tray (this is very important because the area of scan is fixed to cover the frame when well positioned)!

3. Clean drops or marks on the frame.

4. Pour the sample and add tap water until all the perimeter of the frame's step is covered with water. When you pour the sample and add the extra water you can already accelerate the separation of organisms by pouring the sample homogeneously on the tray and by pouring the extra water on conglomerated areas of the tray (i.e., high density of organisms and thus touching) to "dissolve" them.

6. Take 5-15 minutes to properly separate the organisms. Move to the tray the organisms placed on the step of the frame or touching the frame (if not, they will not be considered in the scan as the image is cropped on the borders of the frame). Also, if some organisms are floating, try to sink them with the wood peak (otherwise their size measurements are wrong and their image captions are blurred). If it's difficult to sink the floating organisms and they are very few, the best is to take them out of the image (e.g. placing them on the step of the frame). This step (organisms' separation) is critical to have good data quality. Thus, please pay attention and try that none object is touching another one. Nevertheless, some samples can be difficult to separate. Make a compromise between the time expended separating and the quality of the image. You can, after the process of the image, separate touching objects with the separation tool in zooprocess.

7. Launch Zooprocess, select your project and click on SCAN sample with Zooscan (for archive, no process). Then, follow the instructions on zooprocess window.

If you would like to use our matlab files to read and treat your data "par default" you must write the date of sampling and fraction as follows in this example:

Mc19980530d1 (this is the large size fraction of a sample from the 30th May 1998 at mc station in Naples).
If you do not need to divide your samples in two size classes, please add d1 at the end of the
name anyway. The matlab program will search for it, and if there is not d2 subsample it will work anyway.

8. IMPORTANT! Do not forget to turn on the green light of the ZooScan before launching the scan in Vuescan. WAIT 30 seconds between the preview and pressing scan (if you don't wait, the light-tempo can bug!).

After scanning:

Be very careful writing your sample meta data file. The first time you do a sample in a project you write in all the boxes of the meta file. Then, in the case of a time series of a fixed station, you will need to change the volume sampled (if the net changes) and the subpart (i.e., the fraction that has been scanned…16 if you have split the sample 4 times). Do not leave any boxes empty.

When you have written the sample meta file and pressed OK, Image J has finished. Your sample raw image + log.txt file + meta.txt file will be created in your project folder:
e.g., E:\Zooscan_largeframe2400\Zooscan_scan_raw

The log file gives information on the scanning method (parameters).

The meta file gives information on the sampling method (sampling site, sampling methodology,e.g., net size, volume..etc, and on the sampling preparation for the zooscan, e.g., pre-filtering and splitting rate).

The image will be treated when you process the scan with those of the day ”Convert and process image in batch mode” (during the night). Anyway you have other options as scanning and processing one image or converting and then processing one image.

PROCESSING SAMPLES
Read section 9, pages 30-41, of Zooprocess_manual_v6.16.

(recommended to launch the batch mode during the night)

If it is the first time that you process samples with your project, before doing it maybe you want to arrange the size limits of the organisms you want to measure. You can change the ESD thresholds (equivalent spherical diameter) of particles to be considered in the scan (0.3-4 mm by default which is good for the mesozooplankton bulk).

To change the configuration for processing scans, open ImageJ, click on Z as usual, select your project if you were in a different one the last time you used the program, and click on the option “Edit Configuration File”. All the options will be displayed in a box. To know about these options (which I recommend to not change if you are not confident with the changes) *go to section 12.7, pages 85-88, of Zooprocess_manual_v6.16.*

Open Image J, click on Z icon and select “CONVERT PROCESS IMAGES in batch mode”.

It is recommended to leave the configuration by default. If you tick the option “Save thumbnail images of organisms” the image process will take much longer. It might be faster to extract the vignettes afterwards with the option “Extract vignettes for plankton identifier”; or much faster with the option “Extract vignettes in folder according to prediction” once the objects have been automatically recognized by a created Learning set.

At the end of the sample processing you will obtain a new subfolder named as the sample and placed in Zooscan_scan_work\ of your project. In the subfolder you will have a new file which is the .pid file. The .PID FILE is a single file that concatenates the log.txt, the meta.txt, the processing functions applied and at the end the meas.txt which is the table containing all objects (rows) and their measurements (columns). The measurements that you might use to compute size for data analysis are Area, Major (longest axis of the object) and Minor (minor axis of a perfect ellipse of the same area of the object measured). Other measurements correspond to variables of shape and texture for automatic recognition, and of position in the tray. To see what each variable is, please check the “IJ_variables.txt” file.

CHECKING IMAGE QUALITY AND SEPARATING TOUCHING OBJECTS
Read sections 11 and at least 12.1, pages 59; 70-71, of Zooprocess_manual_v6.16

(recommended to do it every morning on the batch launched during the previous night)

Every morning, after a night of processing in batch, you select on Zooprocess to “CHECK process by viewing segmented images”. You select the first image of the batch you performed the previous night and you pass to the next one when you close it. If you want to finish you press the space bar (if any image stays opened on the screen, select the option “close all opened images” on the bottom of Z menu). The opened images are the _msk1.gif and they allow you to see if the background was well extracted from your image, i.e., no areas with many dots. You can also check on this image if the organisms are not aggregating.

If you have doubts on the quality of your image, to better check if organisms are well separated you can use the “view image with outlines” or also “view vignettes”. These tools allow you to see if more than one object were considered as a single object by the system.

Whenever you are not satisfied with the manual separation that you performed on the scanning tray (many objects touching in the image), you can separate them by drawing lines between the touching objects. Select the option “Separation from B/W msk image”. A line will cut the single object that was actually composed of two organisms (or more organisms, and in that case more lines to be drawn). When you have finished drawing lines, you press cancel and Zooprocess rewrites on your datatable of the .pid

file to correct the modified objects. You can do the same using the "Separation using vignettes" option. We recommend leaving ticked the sort vignettes by decreasing surface as the conglomerates of organisms might create large objects. Then, if the organisms have already been recognized, you can also tick the second option to see vignettes by group (i.e., you see all the copepods to "clean them").

ZOOSCAN PROTOCOL FOR AUTO- & SEMIAUTOMATIC RECOGNITION

Read section 13, pages 91-98, of Zooprocess_manual_v6.16

Creating a first Learning set:

Learning set: Objects arranged in groups; it acts as reference to sort sample objects.
Programs used: ImageJ(Zooprocess)+ Plankton Identifier.

Once you have some samples and you have checked their quality (perhaps you have also applied the separation mask). You are ready to build your first learning set that will help to automatically sort all the objects in the scanned samples.

1. Enter your project folder and go to "Pid_process" and into "Pid_results". COPY the .pid files that you want to use to build your Learning set (a representative subset of all your set, e.g., seasons represented…) and paste them in "Unsorted_vignettes_pid".

2. Open ImageJ and click on Z icon; then choose the menu "Extract vignettes for Plankton Identifier".

3. Leave "extract all vignettes" option. If the resolution of your images is not 2400 (we recommend 2400dpi though), please change Resolution to the one you have used. If not leave it at 2400 par default.

   The Gamma value can increase if you think the contrast of the vignettes is not enough (e.g., 1.2), but we usually leave it at 1.1 par default. Leave by default ticked boxes and press OK.

4. If you don't want to sort all objects (e.g, you prefer to create a learningset with subsets of many samples instead of all the objects of a few samples), you can click on the first option: Extract all vignettes and select "Random extraction of vignettes", (change resolution and gamma if necessary: see point 3) and click OK. Then choose the number of vignettes you would like to extract (e.g., 200 by sample, and so 200 raws of each pid table in the list). Click OK.

5. Your vignettes are now with the related .pid files in "Unsorted_vignettes_pid".

To be organized, please create a new folder inside the “Unsorted_vignettes_pid” and name it as “unsorted_date” to keep track of it. Then move all the vignettes and .pids there.

6. Your learning set will be built sorting all those vignettes in groups defined by you in “Pid process” “sorted vignettes”. Go in sorted vignettes folder and create a new folder (i.e., Learningset_date). Then, COPY all the pid files in this new folder. You need the pid files to build the table constituting the learning set!

7. Close Image J and open Plankton Identifier. Click on Learning. And select “Learningset_date” in sorted vignettes. The new window has the Learningset_date on the right (i.e., sorted thumbs). Click on the folders icon up on he right to create groups (i.e., categories as copepoda, cladocera, etc). Then, on the left (i.e., Unsorted thumbs) select your folder in “unsorted vignettes” (i.e., unsorted_date). Select vignettes (no more than 50 each time as because they are copied and not moved you could have windows memory problems) and move them to the group they belong to.

8. Once you have finished sorting ALL the vignettes, click on the icon on the right bottom, “create learning file”. We recommend naming it as Learningset_date to keep track of it. Save it in the Learningset_date folder (notice that this file is a concatenation of all the files used to build it, and that a last column has been added with the recognition of each object). Job completed, continue sorting….Say No and click on Data Analysis.

9. Usually, this first Learning set is very crowded but not all groups are well represented. What we recommend is to create a subset of this learningset to be created in the “Learning_set” folder of Pid_process. We recommend to randomly sort 200 vignettes of each category. To do so, open Image J, Z and select “create subset of a learning set from identified vignettes (random)” option. Select your source folder which is the complete learning set created in the sorted_vignettes folder, and then you select to save the new subset on the “learning_set” folder. Select 200 vignettes per category, OK. The subset new folder is named by the datehour of creation and number of random vignettes.

10. Close Image J and open Plankton Identifier. Enter the Learning box to create a new learningset .pid concatenation. Select the Learnset1_random200 with its folders and original .pid files and create a learningset.txt. Click on learning and select the date_random_200 folder (it needs to have the .pid files of all the samples included in it to be accepted).

11. Before using the Learningset (subset), you can check its performance by testing the recognition on itself (i.e., cross validation). You click on Data Analysis on PkId. On the learning set box, select as Learning file your Learnset1_random200 in the

“daterandom200”subfolder of “learning_set” folder of pid_process in your project, and on the left bottom of the main box select the method Cross-validation4 (Rndm tree). The method consists on one part (random) of the learning set recognizing another (2 folds), and this is repeated x times (5 trials in this case) to obtain strong statistical values. Once you have unticked some variables and created some other customised variables following what recommended at the IJ_variables.txt file you launch the analysis that will produce a confusion matrix (true classification (rows) versus automatic classification (columns)).

You can click on “Start Analysis”. Include in the name of the analysis the date to keep track. Select to SAVE RESULTS IN “Prediction” folder of PID_process folder of your project.
After analysis is completed, quit data analysis and click on show report, to then select the Analysis_name.txt you have created. There, clicking on cross-validation, you can see true classification (rows) versus automatic classification (columns).
The recall is the % of organisms belonging to a group that were automatically well recognised, whereas the 1-precision is the % of organisms classified by the algorithm as a group that they do not belong to (contamination in a group).

Validating object recognition and improving your Learning set:

1. Now, with the lerningset_date_subset we can recognize samples in the project (if you want to improve your learning set you can select samples that you think could implement it, i.e. other seasons, like summer to add a cladocera folder to the learning set, or special samples with rare groups). The scope is to do a pre-automatic sorting to then validate the recognition by resorting the vignettes manually. At the end of the process, we will have our final datasets, and if we want we can have some new vignettes to implement the Learning set.

We strongly recommend to spend very short time in implementing your learning set and to rather validate all your samples after automatic recognition.

In this procedure, automatic recognition is a mean to sort organisms faster. But human eye is who finally decides the classification.

2. In Plankton Identifier you click on Data Analysis and select on the left top box the Learning set (i.e., Learnset1_random200). Bellow, you select in Pid_results folder the samples .pid than you want to automatically recognize (recommended max 20-30 at a time to not have memory issues). Check the variables that will be used to classify the objects (you should have them as in the doc “IJ_variables.txt”, if it’s the first time you use PkId you will have to select the valiables as the list in the last line of the .txt file).

3. Select the SpvLearning4 (Random forest) method. TICK “save detailed results for each sample” option. A file will be created for each sample (if not it will give you just a concatenation of results in one single file!). SAVE RESULTS, as Analysis_yyyymmdd to keep track of the date, in the “Prediction” subfolder of your project’s Pid_process folder.

Each Analysissamplename_dat1.txt file will be the table of objects in the .pid file with a last added column containing the automatic classification (Prediction) of each object (line). When PkId has finished you can close it.

In Prediction folder you can now find the Analysis_samplename_dat1.txt files of your samples. These are the original .pid files with an added last column in the data table which says the group predicted by your learningset (i.e., copepod).

4. Now, the samples have to be validated to create your final datasets, and if you want you can take some vignettes to implement the learning set (don’t forget to also copy the related .pid files to the learningset folder).

5. Before validation!, COPY the Analysis.txt files of the samples you are interested in from the “Prediction” folder to the root of “Pid_results”.

6. To validate the automatic recognition of the selected samples, we extract the vignettes in the predicted groups: in ImageJ, Z select “extract vignettes in folders according to prediction”(zooprocess will read the last column and sort the objects’ vignettes in the different groups).

The menu of this function allows you the random extraction of vignettes (as is the case of extract vignettes for PkId). You just need to select it in the select method box. This option will be ticked just to implement the Learning set, if not we extract all vignettes to validate the whole sample and obtain the definitive datasets. Leave the rest by default (only if your scans have been done with a resolution other than 2400 you need to change it in the vignettes) and press OK. Leave the name of the folder by default (date_hour_tovalidate).

If any of the samples has already been extracted, zooprocess will tell you. It helps to avoid using a sample twice to create a learning set and also to avoid validating the same sample more than once when you are creating your definitive dataset. The folder_tovalidate with the objects automatically classified is in sorted_vignettes in the Pid_process folder. The .txt used have been copied to Dat1_extracted in Pid_results.

7. Now, validate the samples. That is, check the automatic sorting and correct it when necessary. To accelerate this process, we recommend using XnView (free

software). Instead of copying vignettes to one folder to another as PkId, it moves the vignettes and thus, you can move many more vignettes in one selection and no problems of windows memory are encountered. Yet, check well your work as you move vignettes and you don't have a reference folder as when working with PkId.

8. Open XnView and select the folder you want to validate (i.e., datehour_tovalidate). Go through each of the subfolders (groups) and check the vignettes. If classification is wrong move the vignette to the right folder (sometimes it will be necessary to create new folders that were not present in the learning set to classify your samples).

9. Once you have finished, close XnView and open Image J:Z. Select "Load identifications from sorted vignettes". The folder is in sorted_vignettes and you select the _tovalidate folder. Then, leave ticks par default. You could untick "Process detailed statistics" if the samples you have validated are not to implement the learningset but to create your definitive dataset. It's up to you. A results table is created if you left it ticked. In it you can see the performance of the automatic algorithm (random forest) by calculating the vignettes corrected and thus recall and 1-pred rates as well as the global error rate (N of vignettes moved).

10. The .txt files have been renamed by deleting the Analysis_ beginning of the name (final data name) and copied in the Dat1_validated of Pid_results. These new tables have a new last column which contains the true classification of each object.

11. Now in the folder "datehour_tovalidate" some .txt files have been added. A part from the statistics file, there is a .txt file per sample included in the vignettes' set. The .txt files without the "Analysis_date_" are the ones used to compute size spectra. These files have been automatically copied to Pid_results Dat1_validated. So each time you will perform the automatic recognition of your samples+validation by manual sorting, the new dat1.txt files created will be stored with the previous ones in Pid_results. Once your project is finished, you can take all of these data and compute abundance and size spectra to analyse.

12. To add new sorted vignettes to the old learning set, you just have to copy the vignettes in each group to the same group in the learning set. Do not forget to then add the .pid files of the last samples you have used to the completed learningset (in the main root of the learning set to be able to compute the new learningset table). Then you just have to load the new learning set in PkId Learning and create the learningset.txt.

We recommend that you re-do step 8 of the previous section (you obtain a new learningset with 200 vignettes for each category, but now some rare categories will be fuller and maybe you have been able to create new categories during the process).

While with automatic recognition we try to recognize very fast a few well distinguished groups. When we validate data, we can add new groups or separate vignettes of a group in more specific ones (e.g., copepod transformed in three groups: cop_calanoid, cop_oithona, cop_oncaea). Do not create new subfolders in the original single folder (e.g., calanus and temora folders inside copepod forder NO), but create parallel new folder and then move the corresponding vignettes to those.

Finally, we will treat the data samplename_a_1_dat1.txt (which have the two columns of recognition), stored in Pid_results.

COMPUTING AND ANALYZING SIZE SPECTRA (some tips)

Datasets are in the form of a table in the "dat1.txt" files. We can provide you with matlab scripts to read these data and to compute the size spectra, calculate abundance and to calculate some variables on the spectra. If you do not want to use the matlab scripts, you can do your calculations with excel. You just import your data in excel and select the ";" separator to recognize the columns of the table.

Converting pixels to milimeters

You need the "Resolution" in the [Info] field of the first part of the _dat1.txt. The resolution is used to calculate the dimension of 1 pixel as follows:

Milimeter/1pixel=25.4/Resol*1000)*10^{-3} (if Resol=2400; 10.5833 microns/pixel);

So the values of "Area" column have to be transformed by multiplying it (e.g. for 2400 resolution: Area*10.5833^2 (because the area is in 2 D), and Major=Major*10.5833.

Calculating sampled abundance/m^3 from abundance in the scanning tray

You need to know the volume sampled with the net and the splitting ratio (subpart scanned) to calculate the abundance/m^3. You will calculate the real abundance as follows:

N=Number of raws with same prediction (e.g., in last column written "copepod").
Subpart=splitting ratio.
Vol=net volume.
Townb=number of net tows in a sample.

N of organisms/m^3= N*Subpart/(Vol*Townb);

You can calculate the abundance of each of your groups by just sorting your organisms by their classification in the last column of your data table.

Calculating ESD (equivalent spherical diameter) and Ellipsoidal Biovolume

From the Area, Major and Minor parameters we can compute the Spherical (SBv) and ellipsoidal biovolume (EBv) with the following equations:

$$SBv = {}^{4}/_{3}\,\pi\left(\frac{Area}{\pi}\right)^{3/2}$$

$$EBv = {}^{4}/_{3}\,\pi\left(\frac{Major}{2}\right)\left(\frac{Minor}{2}\right)^{2}$$

Creating the size spectra

Size spectra are just histograms computed on size classes. To create your spectrum you need to define your size classes and then sort the organisms (all, or of each group) by their size (ESD or ellipsoidal volume). You can just sort them (i.e., you add the body volume of each organism to the size class in which this body volume falls).

It is better to have size classes of geometrical-scale (each size class is wider than the previous one) than equal size classes (all size classes have he same width). Larger organisms are less abundant and have larger mean size; therefore we need wider size classes to better represent their size distribution. In particle size-spectrum analysis the use of the octaves-scale (a geometric 2^n series) has been applied since first description of particles-size distribution in the ocean. This scale allows size classes to be defined by the entire base of the logarithmical scale, and the amplitude of the bin is the same as its low limit.

In the matlab scripts we provide you with, size classes have an scale of $k=2^{1/4}$ with the lowest and largest size limit:

smin=0.001; lower limit of the biovolume spectra that will be calculated
smax=10000; upper limit of the biovolume spectra that will be calculated

(in this case 94 size classes to characterize the whole zooplankton spectrum)

Size classes are defined by their average size = (maxLim-minLim)/2.

Once you have computed your spectra, you need to multiply the value in each size class by the conversion factor we used to compute abundance/m^3:

spectrum*Subpart/(Vol*Townb);

At this step, the spectrum units are mm^3/m^3.

If you have more than one image per sample (d1 and d2 fractions). After converting the spectra you can add them together.

Spectra are then normalised by dividing the biovolume in each size class by the width of each size class. Finally, spectra are converted to logarithmical scale to obtain a linear spectrum.

Units are on the *y*-axis in log (mm^3 m^{-3} mm^{-3}) and the *x*-axis shows individual biovolume in log(mm^3).

Computing some parameters to describe the spectra

Spectra can be represented by the **slope** of their linear regression (of the whole spectrum or from their mode):

Y=bx+a

In which b is the slope.

The slope can be computed in the whole spectrum or from the mode of it (size classes with the maximum biovolume). The average mode for a study is computed as the mode frequency of the modes of all your spectra of your project. This mode is due to net sampling underestimation of small organisms of similar ESD or a bit higher than the net mesh.

The Shannon index on the size classes (**size diversity**) can also be used as an indicator:

Vi=vol in each size class;
Vol= Vol total spectrum;
pi=Vi/Vol;

$$H' = \sum_{i}^{S} p_i \log_2 p_i$$

The slope is around –1. The more negative the slope is (and the lower the Sh index is) it indicates that the spectra has a higher proportion of small organisms (and *viceversa*).

You could also calculate an abundance spectrum (number of organisms instead of volume in each size class) and then cut the spectra to observe the abundance changes in the smallest and largest size classes.