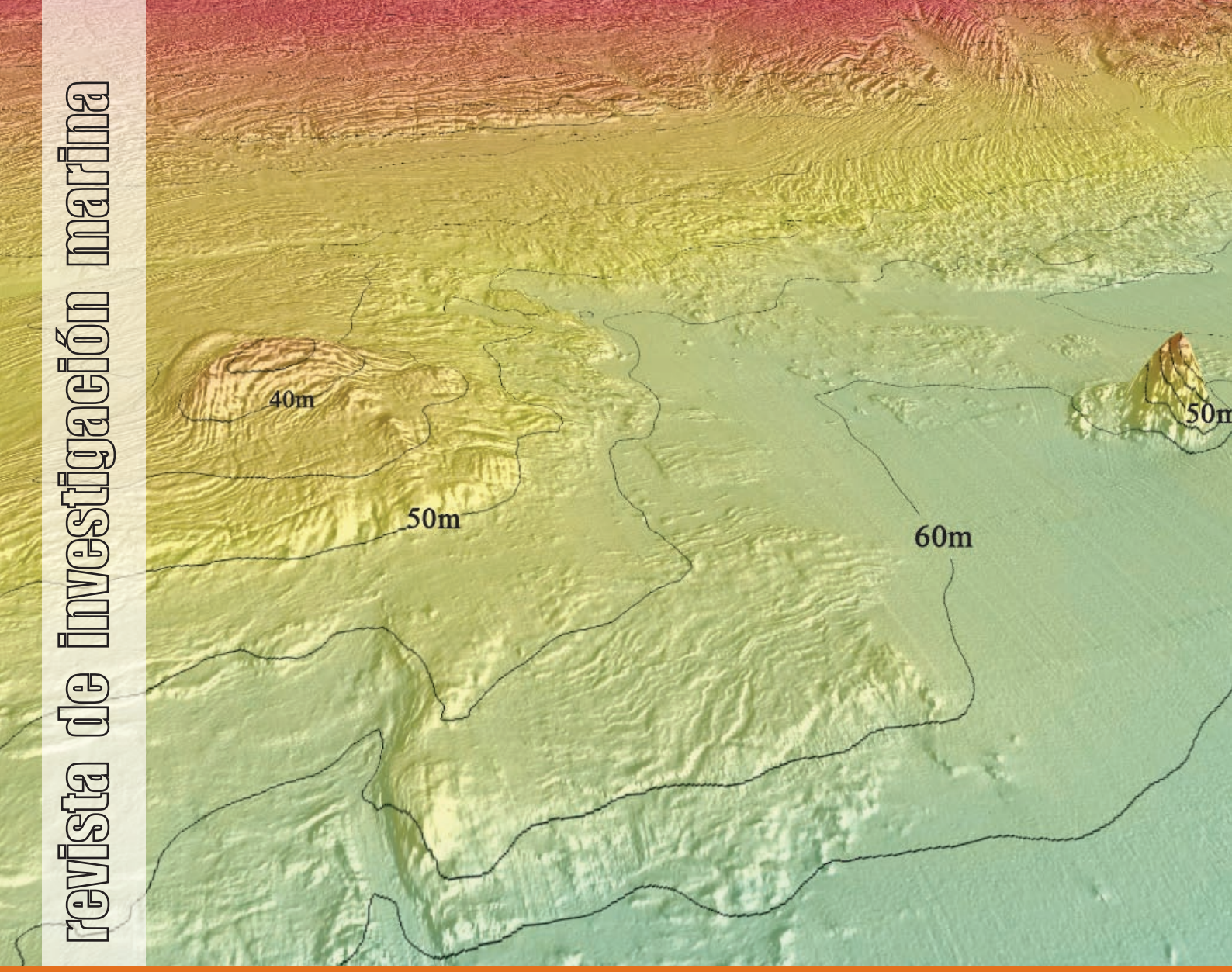
revista de investigación marina



18(2)

Zooplankton Image Analysis Manual: automated identification by means of scanner and digital camera as imaging devices

Eneko Bachiller
Jose Antonio Fernandes

azti tecnalia

Bachiller, E., Fernandes, J.A., 2011. Zooplankton Image Analysis Manual: Automated identification by means of scanner and digital camera as imaging devices. 18(2): 16-37.

La serie *'Revista de Investigación Marina'*, editada por la Unidad de Investigación Marina de Tecnalia, cuenta con el siguiente Comité Editorial:

Editor: Dr. Ángel Borja

Adjunta al Editor: Dña. Mercedes Fernández Monge e Irantzu Zubiaur (coordinación de las publicaciones)

Comité Editorial:
Dr. Lorenzo Motos
Dr. Adolfo Uriarte
Dr. Michael Collins
Dr. Javier Franco
D. Julien Mader
Dña. Marina Santurtun
D. Victoriano Valencia
Dr. Xabier Irigoien
Dra. Arantza Murillas
Dr. Josu Santiago

La '*Revista de Investigación Marina'* de Tecnalia edita y publica investigaciones y datos originales resultado de la Unidad de Investigación Marina de Tecnalia. Las propuestas de publicación deben ser enviadas al siguiente correo electrónico aborja@azti.es. Un comité de selección revisará las propuestas y sugerirá los cambios pertinentes antes de su aceptación definitiva.

Edición: 1.ª Febrero 2011
© AZTI-Tecnalia
ISSN: 1988-818X
Unidad de Investigación Marina
Internet: www.azti.es
Edita: Unidad de Investigación Marina de Tecnalia
Herrera Kaia, Portualdea
20010 Pasaia
Foto portada: © AZTI-Tecnalia

© AZTI-Tecnalia 2011. **Distribución gratuita en formato PDF a través de la web: www.azti.es/RIM**

Zooplankton Image Analysis Manual: automated identification by means of scanner and digital camera as imaging devices

Eneko Bachiller[1*], Jose Antonio Fernandes[1]

Abstract

Rapid development of semi-automated zooplankton counting and classification methods has carried out new chances when defining objectives for plankton distribution studies. Image analysis allows processing many more samples than under microscope classification with less effort and faster, but with lower taxonomical resolution. Although research in this field has been recently focused in scanning devices, different equipment for image capturing such as photographic cameras can offer alternative utilities. In this manual the whole zooplankton sample processing is explained according to laboratory protocols followed in AZTI-Tecnalia, from sample preparation to automated taxonomic identification using scanner and digital camera for digitizing samples and *ZooImage* as software. In addition, a new internal control methodology is proposed in order to obtain a reliable quality check of the whole zooplankton identification analysis procedure: if an error occurs in any step of the procedure, this will be reflected on results.

Key Words

Zooplankton identification, Automated classification, Digital imaging devices, ZooImage, Internal Procedure Quality Check.

Introduction

In case of traditional zooplankton identification methodology, the influence of the expert (Culverhouse *et al.*, 2003; Benfield *et al.*, 2007) and the limited number of samples that can be accurately processed in a cost-effective time and effort (Tang *et al.*, 1998; Grosjean *et al.*, 2004; Boyra *et al.*, 2005; Benfield *et al.*, 2007; Bell and Hopcroft, 2008; Gislason and Silva, 2009) have supposed an increasing interest for develop new identification tools. Hence, the combination of manual counting with new technologies would contribute to a better understanding of the structure and functioning of planktonic ecosystems, as well as to obtain other results such as size and biomass that would not be easily achieved with conventional methods (Huntley and Lopez, 1992; Alcaraz *et al.*, 2003; Grosjean *et al.*, 2004; Culverhouse *et al.*, 2006; Irigoien *et al.*, 2006; Benfield *et al.*, 2007; Gislason and Silva, 2009; MacLeod *et al.*, 2010). Rapid development of semi-automatic zooplankton counting and classification methods has carried out new chances when defining objectives for plankton distribution studies (Benfield *et al.*, 2007; MacLeod *et al.*, 2010).

Automated identification allows increasing spatial and temporal resolution of the study as well as processing more samples with much less effort and reasonably faster. This would also transform alpha taxonomy to a much more accessible, testable and verifiable science (Benfield *et al.*, 2007; MacLeod *et al.*, 2010), with the possibility of saving plankton sample records in a digital format and preventing the loss of information due to both deterioration in the preservative (Ortner *et al.*, 1979; Ortner *et al.*, 1981; Leakey *et al.*, 1994; Alcaraz *et al.*, 2003; Zarauz, 2007) and sample manipulation (Benfield *et al.*, 2007).

The laboratory image capture in this field of research has recently focused in scanning devices. However, there are different image capturing devices such as photographic cameras that can offer alternative utilities.

Objectives

In this manual, the whole sample processing procedure is explained, both for scanner and digital camera as imaging devices, as well as many different experiments made in order to improve the accuracy of obtained results. In addition, a new internal control methodology is proposed in order to detect any error during the sampling or image processing (Harris *et al.*, 2000; Benfield *et al.*, 2007), obtaining reliable and quantifiable quality check of the whole zooplankton identification analysis procedure.

Zooplankton sampling

Samples processed with this methodology come from oceanographic surveys carried out aboard research vessels that cover the southeast of the Bay of Biscay (*R/V Investigador*, *R/V Emma Bardán*...). A vertical plankton haul is made at each sampling station (Figure 1), using a 150 µm or 63 µm PairoVET net (2-CalVET nets, (Smith et al., 1985)). The net is lowered to

[1] *AZTI-Tecnalia; Marine Research Division*
Herrera Kaia, Portualdea z/g; 20110 Pasaia; Spain,
** ebachiller@azti.es*

a maximum depth of 100 m or in case of shallower stations, 5 m above the bottom. Samples are preserved in 4% formaldehyde buffered with sodium tetraborate (Harris et al., 2000), stored in 250 mL jars.

Figure 1. Zooplankton sampling with PairoVET net during *BIOMAN07* survey onboard *R/V Investigador*

Sample preparation and application of Internal Control Methodology

The proposed internal control methodology consists on adding a previously known amount of Control beads into the plankton sample bottle, in order to detect any anomaly during the whole process, since we are expecting to have a defined abundance range of those beads in later obtained subsamples and also in final results.

Those Control beads should have similar behaviour as zooplankton when shaking the sample bottle, in order to avoid any artificial tendency when taking subsamples with both Hensen pipette or Folsom's Plankton Divider. In this protocol, *Amberlite™ XAD-2 Polymeric Adsorbent* (www.supelco.com) resins (Figure 2, Table 1) were selected for that purpose, called here as Amberlite beads.

Taking a volume of 1 mL of wet Amberlite beads (previously filtered through a 500 µm sieve) in a graduated measuring glass pipette, beads were manually counted under microscope. Three manual counting of each of three 0.5 mL replicates were made, defining an abundance of 2756 (St. Dev. ±84) Amberlite beads per

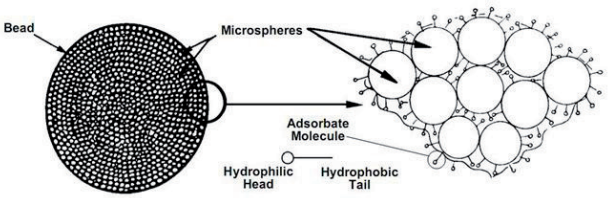


Figure 2. Structure of a Hydrophobic, Macroreticular Amberlite XAD-2 Resin Bead (from *SUPELCO®*)

Table 1. Typical physical properties of Amberlite XAD-2 resin (from *SUPELCO®*)

Appearance:	**Hard, spherical opaque beads**
Solids:	55%
Porosity:	0.41 mL pore mL^{-1} bead
Surface Area (Min.):	300 m^2 g^{-1}
Mean Pore Diameter:	90Å
True Wet Density:	1.02 g mL^{-1}
Skeletal Density:	1.08 g mL^{-1}
Bulk Density:	640 g L^{-1}

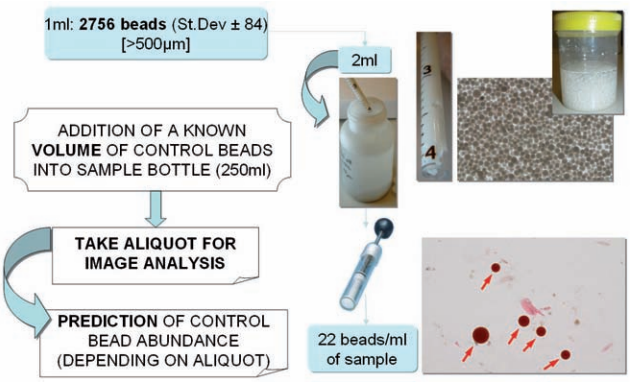


Figure 3. Schematic diagram of the proposed Internal Control Methodology that consists on adding *Amberlite Beads* into zooplankton sample bottle

mL. Subsequently, a volume of 2 mL of wet Amberlite beads was added to a 250 mL zooplankton sample bottle (whole sample), which should result in a concentration of 22 Amberlite beads per mL. A schematic diagram of this procedure is presented in Figure 3.

Subsampling and preparation of aliquots

- Firstly the whole sample volume has to be measured; hence, once Amberlite beads had been added, sample has to be smoothly removed and spilled in a test tube, in order to note the **initial volume (mL)**.
- Then, subsampling can be made (Hensen pipette), as well as many replicates if necessary depending on aim of the study and accuracy needed. **Aliquot volume: 5 mL**
- All subsamples should be **stained** for with 1 ml Eosin (5 g L^{-1}), in order to stain the cell cytoplasm and the muscle protein and so that creating sufficient contrast to be recognized by image

analysis. 0.5 mL Eosin should be added to each sample for 24 hours.

- Subsamples have to be spilled on microtiter polystyrene plates (126 x 84 mm) for later analysis with different methods (a 63 or 150 m sieve will be used for that, depending on mesh size of net used for sampling in each case).
- Some important advices:
  - In order to avoid bubbles **use temperate water (i.e. 30-50 °C) to spill sample on plate!!**
  - Preferably use **new** plates, in order to avoid any scratched or dirty plates
  - Samples should cover the whole surface of the plate, using the minimum amount of water for that.
  - No zooplankton individual should be in contact with plate borders, in order to avoid later aggregates or non useful extracted vignettes; plastic tweezers are used to distribute the zooplankton all over the plate.

Digitizing samples

Scanner: EPSON V750 PRO

Previously prepared zooplankton sample plates can be scanned at 2400 dpi or 4800 dpi resolution using the scanner (Figure 4).

Figure 4. EPSON V750 PRO scanner system

- Place prepared zooplankton sample plates on the scanner (**2 samples per each scanning**).
- Switch on the computer.
- Switch on the EPSON V750 PRO scanner.
- Open *VueScan Professional Edition 8.5.02* software.
- *File → Load Options* (Figure 5).
  - Load predetermined option. In case of this scanner, a file has been prepared fixed to the size of the plankton sample template adapted (e.g. *Bi_plaka.ini*).
- *Output → Load Options* (Figure 6).
  - Check *default folder* and *JPEG file name* sections to ensure the destination and name structure of pictures (e.g. *ECO09_EB_P60_001+.jpg*).

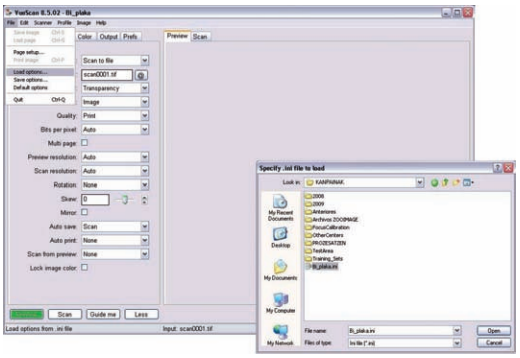


Figure 5. Screenshot of loading options before scanning samples

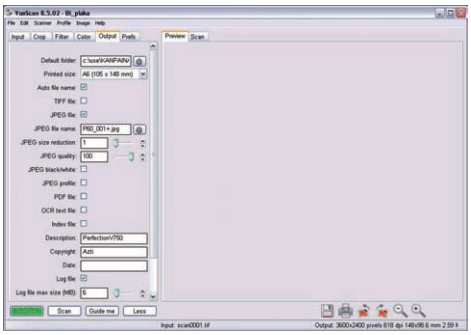


Figure 6. Screenshot of file naming options before scanning samples

- All created files (images) have to be **named exactly as in *ImportTemplate.zie* file**, since *.zim* files that are going to be created later on have to be named equally as well. Use of renamers is optional depending on each study.
- See ANNEX I to see how to name files in a correct way.

Digital Camera: CANON EOS 450D

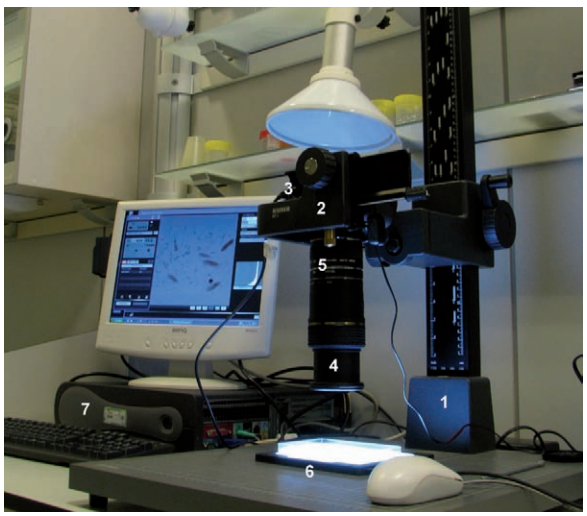


Figure 7. CANON EOS 450D camera system. **(1)** Copy Stand & Tilting Arm (Model Kaiser RS1 5511); **(2)** Micrometric Sliding Plate (Model Manfrotto 454); **(3)** Camera (Model CANON EOS 450D); **(4)** MACRO Lens (Model Tamron SP 90mm F/2.8 Di EOS); **(5)** Extension Tubes (Model KENKO 3 Ring DG P/ Canon EOS, 12+20+36mm); **(6)** Uniform White LED Backlight (Model BIBL-w130/110); **(7)** Connected Computer System.

This system consists on a Copy Stand & Titling Arm (Model Kaiser RS1 5511) and a Micrometric Sliding Plate (Model Manfrotto 454) with a Canon EOS 450D digital camera controlled from the computer (Figure 7). The optic consisted on a Macro Lens (Model Tamron SP 90 mm F/2.8 Di EOS) and Extension Tubes (Model KENKO 3 Ring DG P/Canon EOS, 12+20+36 mm). In addition to that, since a uniform background light is essential for an appropriate later vignette extraction, a White LED Backlight (Model BIBL-w130/110) was also provided. This configuration allows for different resolutions depending on the focus distance and Macro Lens adapted (Table 2). Figure 8 shows the effect of increased resolution in the visual aspect of individual organisms in extracted vignettes.

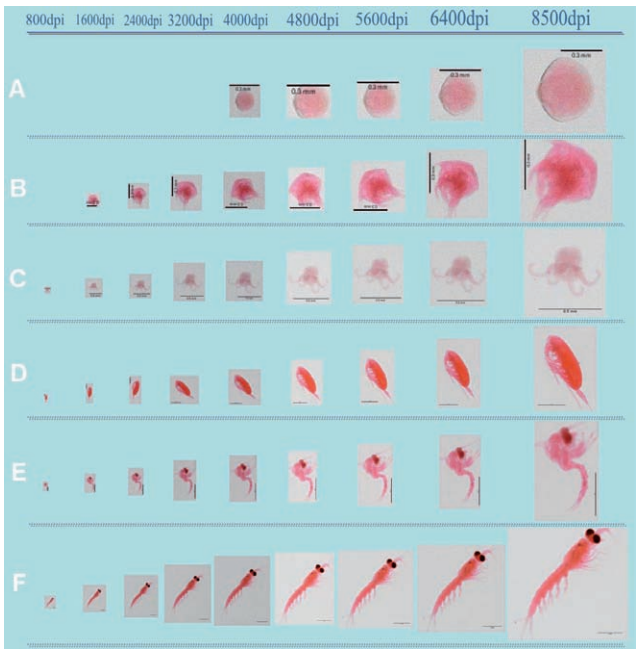


Figure 8. Vignettes show different clarity depending on resolution of images from which they have been extracted. This way the realization of the Training Set can be easier or harder for the expert, depending on size and abundance of vignettes and required accuracy in classification. All this vignettes have been extracted from images taken with the digital camera. **(A)** Bivalve veliger; **(B)** Cirriped nauplius; **(C)** Cephalopoda larva; **(D)** Calanus sp.; **(E)** ZOEA larva; **(F)** Euphausiid.

As an example, this manual is going to consider the option of *8500 dpi resolution* for later explanations:

- Take off the camera lens cover.
- All extension tubes have to be adapted to objective lens (full macro), i.e. 68 mm. (if not, now is time for that).
- Switch the backlight on.
- Switch the camera on.
- Open the *EOS Utility* icon on the desktop.
- Click on the option "*Camera settings / Remote shooting*" (Figure 9). If it is not enabled, try to unplug the camera or turn it off/on with the *EOS Utility* program closed, wait, and open the program again.

Figure 9. Screenshot of *EOS Utility* opening window

- Press "*Remote Live View Shooting*" button in the lower right part of the control window (Figure 10). Real time remote shooting window will open.
- Set camera configuration according to the selected resolution (e.g. 8500 dpi).
  - Set camera position in stand to 37.2 cm (looking from the top of the adapted sliding plate).
  - Diaphragm: F6.3
  - ISO: 1600
  - Obturation Velocity: 1/320 s
  - White Balance: Custom (a photo to illumination alone)

Resolution (dpi)	Camera position in Stand (cm)	Macro lens mm (extension tube)	Magnification Factor	Pixels per mm (±Desv.Est.)	Each photograph Area's width (mm)	Each photograph Area's heigth (mm)	Num. of photos inside plate (NO overlap)	Covered plate area with photos (%)
800	73.8	12	9x	33.66 ± 0.81	126.92	84.61	**1**	**105.61**
1200	60.3	20	12x	47.16 ± 1.16	90.59	60.39	**1**	**53.80**
1600	50.3	20	17x	65.83 ± 0.4	64.89	43.26	**2**	**55.22**
2000	47	36	22x	83.66 ± 0.81	51.06	34.04	**4**	**68.38**
2400	44.6	36	24x	94 ± 0.89	45.45	30.30	**4**	**54.17**
2800	43.7	48	28x	108.33 ± 1.5	39.44	26.29	**8**	**81.57**
3200	42	56	33x	126.83 ± 0.75	33.68	22.46	**9**	**66.95**
3600	40.5	56	37x	145.16 ± 3.43	29.43	19.62	**10**	**56.79**
4000	41	68	40x	150.16 ± 4.16	28.45	18.97	**12**	**63.68**
4400	39.6	68	46x	177 ± 1.67	24.14	16.09	**16**	**61.11**
4800*	**38.9**	**68**	**50x**	**193.66 ± 1.36**	**22.06**	**14.71**	**25**	**79.76**
5200	38.5	68	54x	206.5 ± 1.51	20.69	13.79	**25**	**70.15**
5600	38.1	68	57x	220.83 ± 1.16	19.35	12.90	**25**	**61.34**
6000	37.8	68	61x	235.33 ± 0.81	18.15	12.10	**32**	**69.14**
6200	37.6	68	64x	248.33 ± 1.21	17.20	11.47	**32**	**62.09**
6400	37.5	68	67x	255.83 ± 2.04	16.70	11.13	**36**	**65.82**
6800	37.4	68	69x	269.42 ± 0.90	15.86	10.57	**36**	**59.34**
7000	37.2	68	71x	277.32 ± 0.93	15.40	10.27	**42**	**65.35**
7200	37.3	68	74x	285.22 ± 0.96	14.98	9.99	**42**	**61.78**
7600	37.3	68	77x	301.02 ± 1.02	14.19	9.46	**60**	**79.23**
8500*	**37.2**	**68**	**80x**	**337.57 ± 1.17**	**12.69**	**8.46**	**78**	**82.39**

Table 2. Possible configurations of the Camera System in order to obtain different resolutions in images. In all cases the camera was configured at ISO1600, f/22 at 1/80 s, 4800 dpi and 8500 dpi resolutions are highlighted on table since they have been used for this work.

Figure 10. Screenshot of *Remote Live View Shooting* option in *EOS Utility*

- Photograph the 1 mm (accuracy: 0.01 mm) microrule (*calibration photo*) inside the plate, with the rule on the bottom side (next to the bottom of the plate). **Do not touch the camera** at all after this step, until finish taking all photographs of the plate.
- In case of wanting to check the **real resolution** obtained, go to *Pixel Size* section on ANNEX II.
- Place the 8500 dpi transparency-grid over the backlight illumination.
- Place previously prepared zooplankton sample plate on the backlight, over the prepared transparency grid and below camera lens.
- Set the proper file name format and destination folder for images; all created files (images) have to be **named exactly as in *ImportTemplate.zie* file**, since *.zim* files that are going to be created later on have to be named equally as well. Use of renamers is optional depending on each study.
- See ANNEX I to see how to name files in a correct way.
- Take one photo to each *box* of the grid, moving the plate as smoothly as possible and not touching the camera at all. All photos have to be taken **clicking with the mouse** (*EOS Utility*).

Image Processing. The use of *ZooImage*

Spread sheet preparation

- Prepare an empty folder on the hard disk.
- Open *ZooImage*.
- Select the active directory:
  - *Options → Change active dir…*
  - Select the folder just created.
- Copy the "*SpreadSheet-example*" file in our new folder. It can be downloaded from the *ZooImage* website (www.sciviews.org/zooimage/). It contains:
  - *ImportTemplate.zie*
  - *Zooimage-example.txt*
  - *Zooimage-example.xls*
  - 10 images (examples), to be deleted or replaced with our previous scanned images.
- Open the *Excel* file (e.g. *ZooImage-template.xls*) with *Microsoft Excel*. Now it has to be modified depending on our samples.
  - ***Some important advices:***
  - Never change the order or name of original columns in blue and orange!!
  - New columns can be added at the end, but be careful (previous point).
  - Name images as *p-0001.jpg, instead of p-1.jpg* (use automatic renamers).
  - Use the current date format (yyyy-mm-dd): 2005-12-22
  - Use "." notation for decimal number.
  - Remove stations where there is no image from the Spread Sheet.
  - Make sure there is no empty line at the end of the file!!
- Save the file as *Text (tab delimited)* i.e.: ***.txt*** in the same folder where the original *.xls* file is located. It is possible to have to replace the existing *.txt* file by the new.
- Finally, the working folder should have:
- *Zooimage-example.xls* (keep *Excel*'s column format untouched!)
- *Zooimage-example.txt* (ensure that there is no empty lines at the end!)
- *ImportTemplate.zie* (check it with a plain text editor for possible modifications)
- All digitized images, with optimized format names, i.e. **same name as in *.txt*!!** (ANNEX I).

Sample processing

Importing images

- Open *ZooImage*.
- Click on *Analyze → Import images…*
- Open the *.txt* file created from the original *Excel* file (e.g. *Zooimage-template.txt*).
- In this step the computer can be left **unattended** (Figure 11) since any problem during the processing will be reported as an informative message.

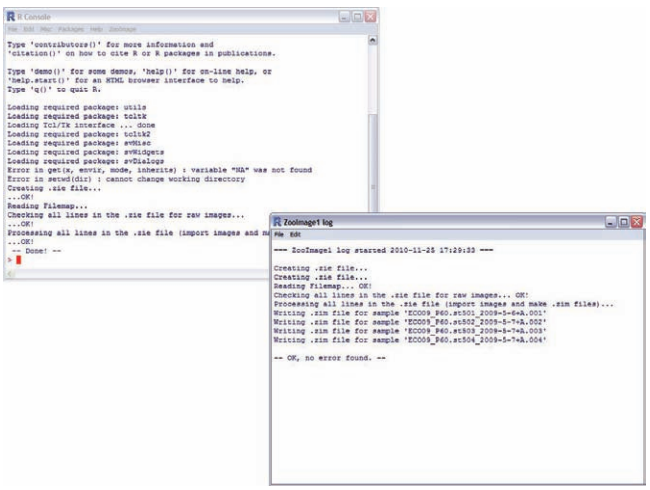


Figure 11. Screenshot of *R console* and *ZooImage log* windows while importing images

- At the end, some files and new folders should be obtained:
  - [_raw]
    - *.zie* files: original *ImportTemplate.zie* and *Import-Zooimage-Template.zie* (with additional data).
    - *.xls* file
    - *.txt* file
    - . *jpg* files (or *.tif*): original images (ANNEX I)
    - Example 1: **one image** per sample or station
    - *ECO09_EB_P60_001.jpg, ECO09_EB_P60_002.jpg...*
    - Example 2: **many images** per sample or station
    - ECO09_EB_P60_001.jpg, ECO09_EB_P60_002.jpg...

    NOTE: Although names are equal in both cases, original images will be renamed in a different way after being processed (depending on aliquot or replicate presence). Depending on *ZooImage* version the extension of scanned images have to be changed manually to "**.jpg**" in this folder.
  - [_work]
    - Images renamed as defined in *.txt* file (at "*Sample*" column) (ANNEX I).
    - If original images (in *_raw* folder) are...
      - Example 1: **one image** per sample or station*ECO09_EB_P60_001.jpg, ECO09_EB_P60_002.jpg...*
      - Example 2: **many images** per sample or station

      ECO09_EB_P60_001.jpg, ECO09_EB_P60_002.jpg...

    and the name wanted for those images on *Excel* or in *.txt* ("*Sample*" column)...
    - Example 1: **one image** per sample or station ECO09_P60_2009-5-6_501+A, ECO09_P60_2009-5-7_502+A....
    - Example 2: **many images** per sample or station ECO09_P60_2009-5-6_501+A, ECO09_P60_2009-5-7_502+A....

      then image names in "*_work*" folder will be:
    - Example 1: **one image** per sample or station ECO09_P60_2009-5-6_501+A.1.jpg, ECO09_P60_2009-5-6_502+A.1.jpg...
    - Example 2: **many images** per sample or station ECO09_P60_2009-5-6_501+A.1.jpg, ECO09_P60_2009-5-7_501+A.2.jpg...

    .zim files, one per station (with no particle information, just data from *Excel*) (ANNEX I)
    - Example 1: **one image** per sample or station ECO09_P60_2009-5-6_501+A.zim, ECO09_P60_2009-5-6_502+A.zim...
    - Example 2: **many images** per sample or station ECO09_P60_2009-5-6_501+A.zim, ECO09_P60_2009-5-6_502+A.zim...
- Before going on the next step, all renamed images (i.e. from *_work* folder) have to be replaced to be processed with their corresponding *.zim* files (named equally) together in the *active directory*, out of "*_raw*" or "*_work*" folders.

Processing images

- Click on *Analyze* → *Process images…*
- *ImageJ* will open.
- Click on *Plugins* → *ZooPhytoImage* → [filter] (f.ex.: *Scanner4800_Colour*)

  NOTE : The wanted filter can be previously compiled using *ImageJCompile* shortcut (see *Filter Compilation* section on ANNEX III).
- Since there is one *.zim* file for each image (or station just in case of having many images of the same plate), selecting only the first *.zim* file is enough to process **all** images of our active folder. **It is essential to have *.zim* files named exactly as images**.
- *"ZooImage1 Image Processor"* window will open. Activate the following options:
  - *Process all items in this directory* (all images having associated *.zim* files)
  - *Analyze particles* (measurements of particles after image processing)
  - *Make vignettes* (to extract small images of each particle)
  - *Sharpen vignettes* (to apply a *sharpen* filter on vignettes to enhance quality)
- *ImageJ* will open a "*Log*" window where it will report its activity.
- The vignette extraction process can be followed on the screen (Figure 12).

  NOTE: This process will **slow down** the entire computer for quite a **long time** so that it is not recommended to use any other application while this process is executing.

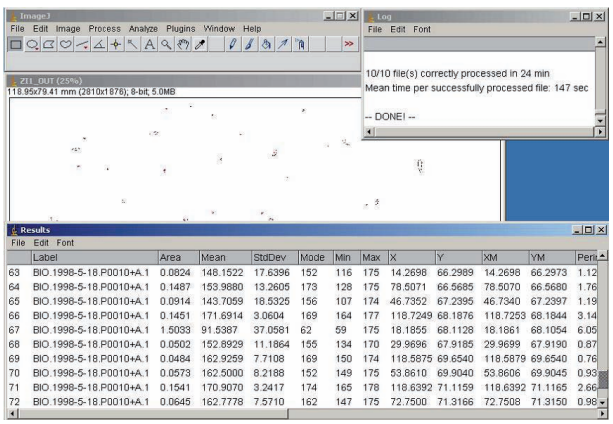


Figure 12. Aspect of the screen while processing images (i.e. vignette extraction and measurement)

- At the end, the working folder should have:
  - [_raw]
    - *.zie* files: original *ImportTemplate.zie* and *Import-Zooimage-Template.zie* with additional data).
    - *.xls* file
    - *.txt* file
    - . *jpg* files (or *.tif*): original images
    - *.jpg* files renamed as *.zim* files: those that have been used to be processed

    NOTE: If any error has occurred during the process, corresponding renamed image will be **out of** the *_raw* folder, as well as its corresponding *.zim* file.

- [_work]
  - *_dat1.zim* files: these files contain general metadata filled for the importation step but also metadata about processing and all measurements done on each particle.
  - Image components:
  - *_msk1.gif*
  - *_vis1.gif*

  NOTE: These files are useful to check how well the selected filter works, overlapping images on *Paint* (see *Filter Compilation* section on ANNEX III).
- *.zim* files: original files (with no particle information, just data from *Excel*)
- One directory by sample (or station) analysed. This directory contains all vignettes, and their *_dat1.zim* files associated (same files as those in *_work* folder).

Creating ZID files

- Click on *Analyze → Make .zid files…*
- Click on *OK*.
- Select the active directory where all our created *.zim* files and other folders (i.e. *_raw* and *_work*, together with folders with vignettes) are located.
- Ensure that there is **no error** reported once compression is done (Figure 13).

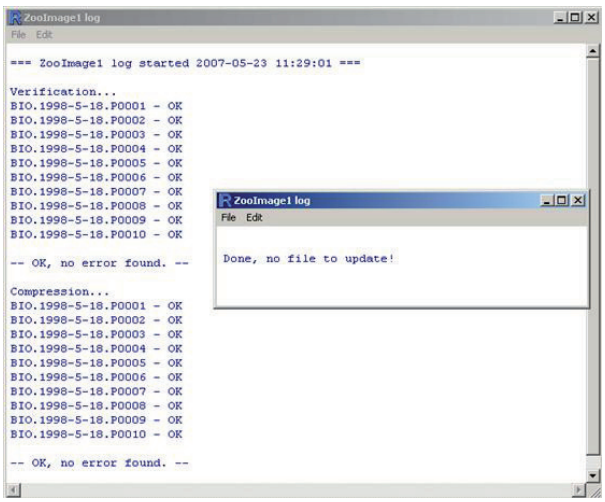


Figure 13. Aspect of the screen while creating ZID files

- At the end, the working folder should have:
- [_raw]
  - *.zie* files: original *ImportTemplate.zie* and *Import-Zooimage-Template.zie* with additional data).
  - *.xls* file
  - *.txt* file
  - . *jpg* files (or *.tif*): original images
  - *.jpg* files renamed as *.zim* files: those that have been used to be processed

  NOTE: [_work] folder usually disappears in this step.
- *.zid* folders: one per sample (or station)
  - All extracted vignettes (*.jpg*)
  - *_dat1.zim* files (with vignette metadata)
  - *_dat1.RData*

  NOTE: These files have to be added in the corresponding folder when improving the Training Set with vignettes from other sources (see *Improving Classifier* section on ANNEX IV).

Before going on the next step, it is advisable to save *.zid* files in corresponding collection folders and to make backups (Figure 14).

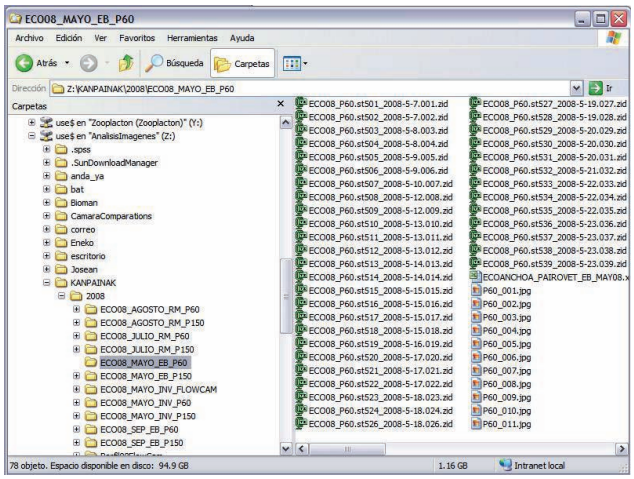


Figure 14. ZID files and original images have to be saved for possible future reprocessing work

What are ZID files?

.zid files are a special kind of zipped archives that contain all that *ZooImage* needs to work with one sample: the *_dat1.zim* files, all vignettes, and a *dat1.Rdata* (compilation of all the data in R format). Therefore *.zid* files can be easily inspected with compression programs (f.ex.: *WinZip, WinRAR...*).

Making the Training Set

- Copy the *.zic* file of corresponding survey (e.g. *Bioman.zic*) into our main folder (*active directory*). A *.zic* file for each survey or period is available in lab PC.

  NOTE: Taxonomic groups of this file can be easily changed and/or added with a plain text editor. Nevertheless, this is not so important since groups can be directly changed and/or created while making the Training Set.
- Click on *Analyze → Make training set…*
- "*Select a .zic file*" window will open.
- Select our *.zic* file (f.ex.: *Bioman.zic*).
- Select the folder in which the training set has to be placed.
- Give a name to the folder of the training set; by defect: "*_train*".
- Select the *.zid* files from which vignettes are going to be used for making the training set (it is not necessary to use all processed samples).
- *XnView* program will open.
  - "_" named folder contains all the images (vignettes) that are going to be taken into account for the realization of the training set. All vignettes are located here at the beginning.
  - Move vignettes to the corresponding folder.

NOTE: If a **new group** is found and no folder has been included in *.zic* file for that, create directly a new folder with *XnView* or *Explorer* (Figure 15).

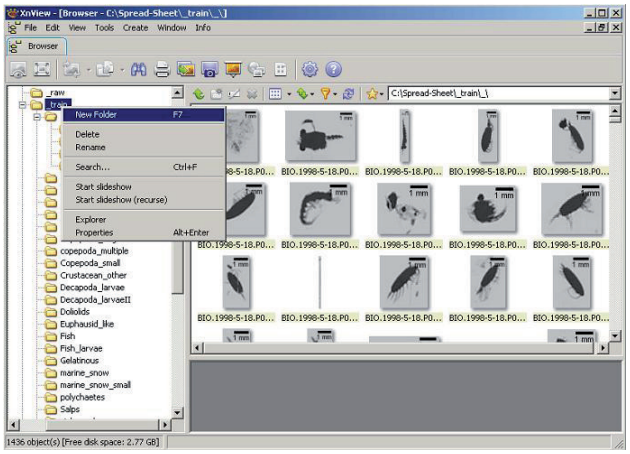


Figure 15. If a new group is found when making the training set, create a new folder within "*_train*" folder.

- At the end, the working folder should have:
- As many folders as taxonomic groups and artefacts with manually classified vignettes inside (Figure 16).

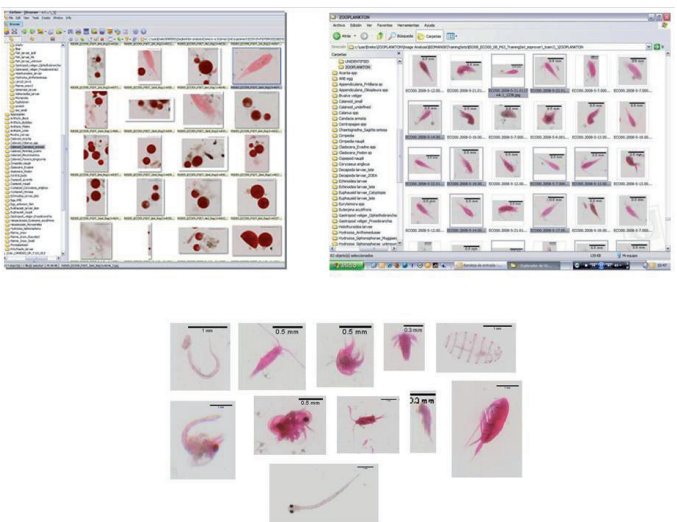


Figure 16. Training Set consists on moving selected vignettes into corresponding group folders in order to train the computer for later automated identification.

- All *_dat1.RData* files from original source of training set images (same as in *.zid*s). These files have to be added in the corresponding folder when improving the Training Set with vignettes from other sources (see *Improving Classifier* section on ANNEX IV).

  NOTE: *Random Forest* classifier can not have empty folders in the training set, neither only one item (i.e. at least two particles are necessary to be considered as valid group).

Reading the Training Set

- Click on *Analyze → Read training set…*
- Select the folder where our training set was created (e.g. "*_train*").
- Give a name to the object that will be created in *R*; by defect: "*ZItrain*"
- Statistics of the Zooplankton classes will be shown in *R console* (Figure 17).

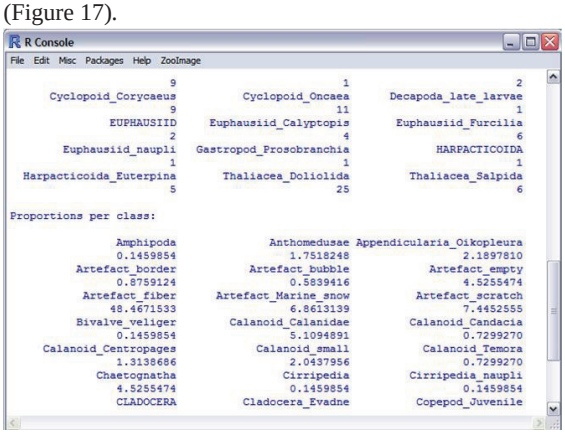


Figure 17. Aspect of the *R Console* window when reading the Training Set.

- Click on *Objects → Save*

  NOTE: It is advisable to save the training set in the corresponding folder as an object in *R*. This way, next time opening *ZooImage* the image analysis procedure can be started from this point just dragging the file into *ZooImage* window, or by clicking on *Objects → Load.*

Making the Classifier

- Click on *Analyze → Make classifier…*
- It is recommended to use "*Random Forest*" option to make the classifier.
- Choose the object created on the last step; by defect: "*ZItrain*".
- Give a name to the object that will be created in *R*; by defect: "*ZIclass*".
- Results of the classifier (accuracy) will be shown (Figure 18).

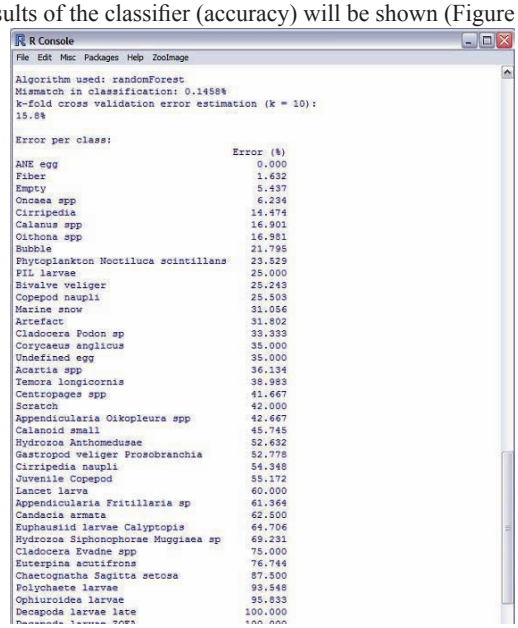


Figure 18. Results of the classifier are presented in *R Console* window.

- Click on *Objects* → *Save*

  NOTE: It is advisable to save the training set object in the corresponding folder as an object in *R*. This way, next time opening *ZooImage* the image analysis procedure can be started from this point just dragging the file into *ZooImage* window, or by clicking on *Objects* → *Load.*

Analyzing Classifier

This is a *Confusion Matrix* to evaluate how good the classifier is. If the error is big or there is a lot of confusion, the training set has to be improved or remade.

- Click on *Analyze* → *Analyze classifier…*
- The diagonals of the *Confusion Matrix* (based on 10 cross fold validation) are the instances well classified (Figure 19).

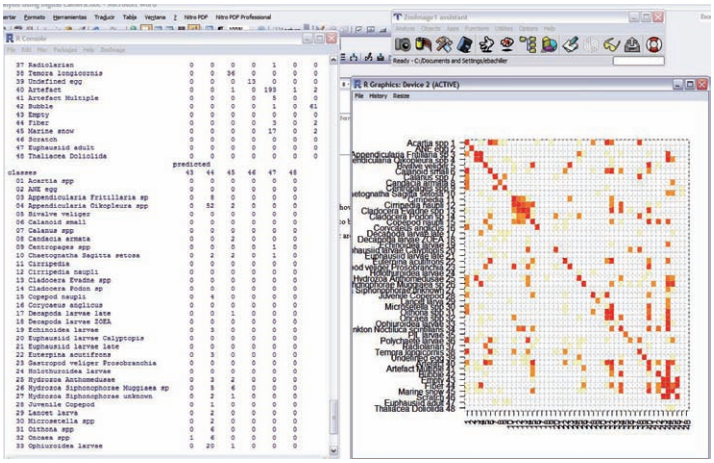


Figure 19. Analyzing classifier. The *Confusion Matrix* helps identifying confusing groups.

- In *Cross Validation Confusion Matrix...* (see *Interpretation of a classifier analysis* section on ANNEX IV):

  Y axis: REAL

  X axis: ESTIMATED (predicted)

  NOTE: Classifier accuracy can be improved with different methodologies (see *Improving the Classifier* section on ANNEX IV).

Treatment and interpretation of results

Processing samples

The following files will be needed to continue with result extraction:

- *Training Set* file (e.g. *ZItrain.RData*)
- *Classifier* file (e.g. *ZIClass.RData*)
- *ZIRes.r*: this file has been modified in order to use the **minor diameter** of particles for their classification. Biomass conversion formula has been used to calculate biomass from images (Alcaraz *et al.*, 2003).

  NOTE: In case of needing results by **equivalent diameter (ECD)** instead of minor diameter:

  - Open *ZIRes.r* file with *TinnR* (Figure 20).

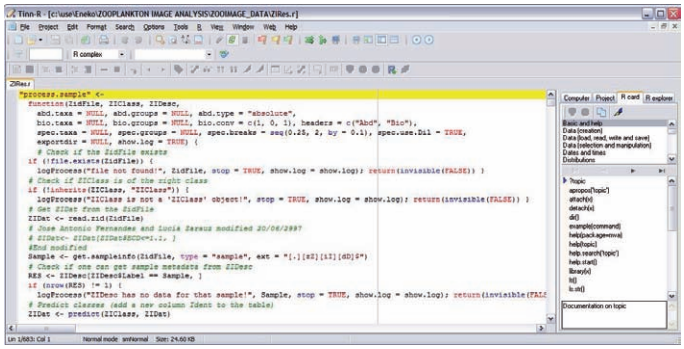


Figure 20. Biomass conversion can be based on minor diameter or equivalent diameter of particles, depending on written code in *ZIRes.r* file.

  - Change symbol "#" before each comment of "ECD" and place it before comments of "MINOR". Use the "*find"* tool to identify all required comments that have to be changed.
  - "#" before each line supposes that it is **not** going to be taken into account. Hence, all sentences related with ECD (or MINOR just in case) should be left activated (i.e. with no "#" symbol before).
- *.zis* file (e.g. *DescripcionAbreviada.zis*): take a template from *ZooImage* website (or from another previously processed survey folder) and modify it with a plain text editor:
  - Open our *Excel* file (e.g. *ZooImage-template.xls*)
  - Save *forZis* sheet data as *.txt* (f.ex.: *ForZis.txt*)
  - Copy all data from *.txt* file and paste it at the end of *.zis* file (f.ex.: *DescripcionAbreviada.zis*). Data such as SST, Salinity, Chlorophyll, etc... can be also added (Figure 21), but make sure that there is no empty line at the end!!

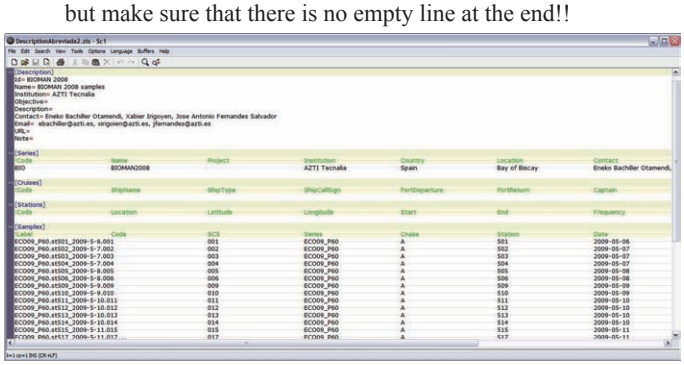


Figure 21. Aspect of a *.zis* file. Additional information of each station can be added here.

  NOTE: *Label* column names should be named exactly as .zid files

  - Save the *.zis* file.
- *Conversion.txt* file: take a template from *ZooImage* website (or from another previously processed survey folder) and modify it with a plain text editor. This file defines which taxonomic groups are adding biomass to our sample and which are non-biological groups.
- *.zid* files of corresponding stations from which results will be

extracted.

- Once all these files have been collected in the main working folder, close all tabs and programs in Windows.
- Open *ZooImage*.
- Charge the Training Set object (e.g. *ZItrain.RData*) in *R*, dragging the file into *R* window, or by clicking on *Object* → *Load*.
- Charge the classifier object (e.g. *ZIClass.Rdata*) in *R*.
- Charge *ZIRes.r* file in *R*.
- Click on *Analyze* → *Process Samples...*
- Select the source of *.zis* file.
- Activate the option to "*Save individual calculations*".
- Select our previously charged Training Set (e.g. *ZItrain*).
- Select our previously charged Classifier (e.g. *ZIClass*).
- Define the **size range interval** in which results have to be received at the end:
  - ***63 µm*** *mesh size samples:* 0.15-15 mm; by 0.2 (**minor diameter* or ECD**).
  - ***150 µm*** *mesh size samples:* 0.20-15 mm; by 0.2 (**minor diameter* or ECD**).
  - ***250 µm*** *mesh size samples:* 0.25-15 mm; by 0.2 (**minor diameter* or ECD**).

  * Defined on *ZIRes.r* file (this could be changed in order to consider ECD instead of minor diameter)
- *ZooImage* will ask to name result object in *R* with a name; by defect: *ZIres.*
- At the end of the process and in order to obtain results in appropriate format for *Excel*, created result objects can be converted to an executable file format.

  Write the following sentences in *R Console* window:

  write.table(ZIres,"Results.csv",sep=",",row.names=FALSE)

  where "*ZIres"* is our created result *R* object and *"Results.csv"*, the output file.

  write.table(attr(ZIres,"spectrum"),"SizeResults.csv",sep=",",row.names=FALSE)

  where "*ZIres"* is our created result *R* object and *"SizeResults.csv*", the output file.

Results

- Two types of files will be obtained from sample processing:
  - *.csv* files (e.g. *Results.csv, SizeResults.csv*)
  - *.txt* files: one file per each processed station, with all individual particle information in it (e.g. *ECO09_P60_2009-5-6_501.txt).*
- ABUNDANCES are in **number of species per m^3**.
  - Abundance per species.
  - Abundance per particle size (minor diameter –by defect– or ECD, depending on what have been previously defined in *ZIRes.r* file).
  - Abundance per species and particle size (minor diameter –by defect– or ECD).
- BIOMASSES are in **mg per m^3.**
  - Biomass per species.
  - Biomass per particle size (minor diameter –by defect– or ECD, depending on what have been previously defined in *ZIRes.r* file).
  - Biomass per species and particle size (minor diameter –by defect– or ECD).
- INDIVIDUAL BIOMASS of particles is in **µg**.
- PARTICLE-SIZE intervals are by defect: **(x_1,x_2]**

Other applications and future perspectives

Development of new imaging systems has been reasonably well funded and both hardware (Wiebe and Benfield, 2003; Culverhouse *et al.*, 2006; Benfield *et al.*, 2007; Schultes and Lopes, 2009) and software (Fernandes *et al.*, 2009; Lehette and Hernández-León, 2009; Fernandes *et al.*, 2010; Gorsky *et al.*, 2010) are continuously being optimized, allowing the sampling of wider distribution areas with less effort and in less time (Gaston and O'Neill, 2004; Benfield *et al.*, 2007; MacLeod *et al.*, 2010). Long-term support of the software accepted by the community, the availability of information systems and networking for the exchange of data and information (Culverhouse *et al.*, 2006; Morales, 2008), such as participation and international collaboration among researchers from diverse academic fields –e.g. Research on Automated Plankton Identification (RAPID) initiative and the Automatic Visual Plankton Identification working group of the Scientific Committee on Oceanic Research (SCOR)– (Benfield *et al.*, 2007) would consolidate this research field allowing to obtain results never expected only with traditional methods. The use of all collected data in different modelling research would also open new fields for further research.

Nevertheless, there is a new challenge that should be considered for further development in high resolution image analysis: *in situ* real-time observation. Despite of some imaging systems are better only for a defined kind of plankton, and still present some limitations (limited volume of sampled water, etc...), *in situ* imaging instrumentation is evolving rapidly (Wiebe and Benfield, 2003; Davis *et al.*, 2004; Remsen *et al.*, 2004; Ashjian *et al.*, 2005; Davis *et al.*, 2005; Culverhouse *et al.*, 2006; Cowen and Guigand, 2008; Schultes and Lopes, 2009; Gorsky *et al.*, 2010; Picheral *et al.*, 2010). However, the high cost (Gaston and O'Neill, 2004) of modern *in situ* instrumentation (such as high resolution 3D imaging systems, ISIIS, Underwater Vision Profiler...) make improbable the replacement of later laboratory analyses by image analysis. In addition, taxonomic accuracy obtained does not seem to reach the same level as obtained processing images in the laboratory with the camera together with manual identification, moreover in a cost effective way.

However, vessels use to stop over for provisions at least once during the survey, hence if some samples would be sent to laboratory for manual identification, results could be combined with those obtained from image analysis in a really short time; moreover and unlike with the scanner, if the digital camera methodology could be effectively applied aboard vessel together with a roll reduction structure, time lag between automated analysis and manual identification would be negligible and taxonomic accuracy, as

high as needed depending on the aim of the project.

On the other hand, digital camera can be useful also for other kind of studies (Figure 22), such as taxonomic identification of stomach contents (records of prey images to compare with other trophic studies or even for later classification, as well as photos of remaining otoliths of preys that have been digested), otolith size structure studies or a target ichthyoplanktonic group counting with

Figure 22. Digital camera offers other possibilities such as taking images of stomach contents, otoliths, etc.

image analysis.

Acknowledgements

We are grateful to AZTI-Tecnalia analysts for their support during the survey such as the sampling process in lab; their effort and interest has been essential for this contribution. Very special thanks are due to P. Grosjean and K. Denis (University of Mons, Belgium) for their help and advice on image analysis, to D. Conway (Plymouth Marine Laboratory, UK) for his help with taxonomic identification of vignettes and to M. Iglesias (scholarship holder in AZTI-Tecnalia during summer of 2009) for her effort with camera configuration experiments. We would also thank the anonymous reviewers for their useful comments. This research was funded by the Department of Agriculture, Fisheries and Food of the Basque Country Government through Ecoanchoa project. E. Bachiller is supported by a doctoral fellowship from the Iñaki Goenaga Zentru Teknologikoen Fundazioa (IG-ZTF).

This is contribution number 525 from AZTI-Tecnalia (Marine Research Division).

References

Alcaraz, M., Saiz, E., Calbet, A., Trepat, I. & Broglio, E. (2003). Estimating zooplankton biomass through image analysis. *Marine Biology*, **143**: 307-315

Ashjian, C.J., Davis, C.S., Gallager, S.M. & Alatalo, P. (2005). Characterization of the zooplankton community, size composition, and distribution in relation to hydrography in the Japan/East Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, **52**: 1363-1392

Bell, J.L. & Hopcroft, R.R. (2008). Assessment of ZooImage as a tool for the classification of zooplankton. *Journal of Plankton Research*, **30**: 1351-1367

Benfield, M.C., Grosjean, P., Culverhouse, P.F., Irigoien, X., Sieracki, M.E., Lopez-Urrutia, A., Dam, H.G., Hu, Q., Davis, C.S. & Hansen, A. (2007). RAPID: research on automated plankton identification. *Oceanography*, **20**: 172-187

Boyra, G., Irigoien, X., Aristegieta, A. & Arregi, I. (2005). Plankton Visual Analyser GLOBEC Norway Annual Science Meeting 2005, pp 9

Cowen, R.K. & Guigand, C.M. (2008). In situ Ichthyoplankton Imaging System (ISIIS): system design and preliminary results. *Limnology and Oceanography: Methods*, **6**: 126-132

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V. & González-Gil, S. (2003). Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Marine Ecololy Progress Series*, **247**: 17-25

Culverhouse, P.F., Williams, R., Benfield, M., Flood, P.R., Sell, A.F., Mazzocchi, M.G., Buttino, I. & Sieracki, M. (2006). Automatic image analysis of plankton: future perspectives. *Marine Ecololy Progress Series*, **312**: 297-309

Davis, C.S., Hu, Q., Gallager, S.M., Tang, X. & Ashjian, C.J. (2004). Real-time observation of taxa-specific plankton distributions: an optical sampling method. *Marine Ecololy Progress Series*, **284**: 77-96

Davis, C.S., Thwaites, F.T., Gallager, S.M. & Hu, Q. (2005). A three-axis fast-tow digital Video Plankton Recorder for rapid surveys of plankton taxa and hydrography. *Limnol. Oceanogr.: Methods*, **3**: 59-74

Fernandes, J.A., Irigoien, X., Boyra, G., Lozano, J.A. & Inza, I. (2009). Optimizing the number of classes in automated zooplankton classification. *Journal of Plankton Research*, **31**: 19-29

Fernandes, J.A., Irigoien, X., Goikoetxea, N., Lozano, J.A., Inza, I., Pérez, A. & Bode, A. (2010). Fish recruitment prediction, using robust supervised classification methods. *Ecological Modelling*, **221**: 338-352

Gaston, K.J. & O'Neill, M.A. (2004). Automated species identification: why not? *Philosophical Transactions B*, **359**: 655-667

Gislason, A. & Silva, T. (2009). Comparison between automated analysis of zooplankton using ZooImage and traditional methodology. *Journal of Plankton Research*, **31**: 1505-1516

Gorsky, G., Ohman, M.D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.B., Cawood, A., Pesant, S., García-Comas, C. & Prejger, F. (2010). Digital zooplankton image analysis using the ZooScan integrated system. *Journal of Plankton Research*, **32**: 285-303

Grosjean, P., Picheral, M., Warembourg, C. & Gorsky, G. (2004). Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES Journal of Marine Science*, **61**: 518-525

Harris, R.P., Wiebe, P.H., Lenz, J., Skjoldal, H.R. & Huntley, M. (2000). *Zooplankton methodology manual*. Academic Press, London. 684 pp

Huntley, M.E. & Lopez, M.D.G. (1992). Temperature-dependent production of marine copepods: a global synthesis. *American Naturalist*, **140**: 201-242

Irigoien, X., Grosjean, P. & Urrutia, A.L. (2006). Image analysis to count and identify zooplankton, Report of a GLOBEC/SPACC workshop on image analysis to count and identify zooplankton, November 2005, San Sebastian, Spain. 21 pp

Leakey, R.J.G., Burkill, P.H. & Sleigh, M.A. (1994). A comparison of fixatives for the estimation of abundance and biovolume of marine planktonic ciliate populations. *Journal of Plankton Research*, **16**: 375

Lehette, P. & Hernández-León, S. (2009). Zooplankton biomass estimation from digitized images: a comparison between subtropical and Antarctic organisms. *Limnol. Oceanogr. Methods*, **7**: 304-308

MacLeod, N., Benfield, M. & Culverhouse, P. (2010). Time to automate identification. *Nature*, **467**: 154-155

Morales, C.E. (2008). Plankton monitoring and analysis in the oceans: capacity building requirements and initiatives in Latin-America. *Revista de biología marina y oceanografía*, **43**: 425-440

Ortner, P.B., Cummings, S.R., Aftring, R.P. & Edgerton, H.E. (1979). Silhouette photography of oceanic zooplankton. *Nature*, **277**: 50-51

Ortner, P.B., Hill, L.C. & Edgerton, H.E. (1981). In-situ silhouette photography of Gulf Stream zooplankton. *Deep Sea Research Part A. Oceanographic Research Papers*, **28**: 1569-1576

Picheral, M., Guidi, L., Stemmann, L., Karl, D.M., Iddaoud, G. & Gorsky, G. (2010). The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnol. Oceanogr. Methods*, **8**: 462-473

Remsen, A., Hopkins, T.L. & Samson, S. (2004). What you see is not what you catch: a comparison of concurrently collected net, Optical Plankton Counter, and Shadowed Image Particle Profiling Evaluation Recorder data from the northeast Gulf of Mexico. *Deep Sea Research*

Part I: Oceanographic Research Papers, **51**: 129-151

Schultes, S. & Lopes, R.M. (2009). Laser Optical Plankton Counter and Zooscan intercomparison in tropical and subtropical marine ecosystems. *Limnol. Oceanogr.: Methods*, **7**: 771-784

Smith, P.E., Flerx, W. & Hewitt, R.H. (1985). *The CalCOFI Vertical Egg Tow (CalVET) Net. NOAA Technical Report NMFS 36. In: Lasker, R. (Ed.), An egg production method for estimating spawning biomass of pelagic fish: Application to the northern anchovy, Engraulis mordax.* US Dep Commer, Washington DC.

Tang, X., Stewart, W.K., Huang, H., Gallager, S.M., Davis, C.S., Vincent, L. & Marra, M. (1998). Automatic plankton image recognition. *Artificial Intelligence Review*, **12**: 177-199

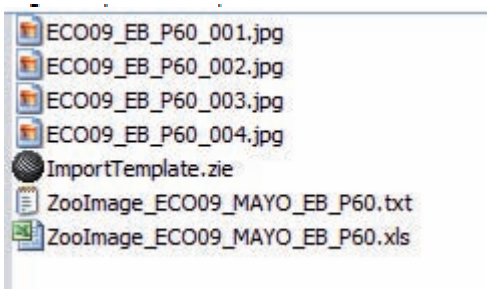
Wiebe, P.H. & Benfield, M.C. (2003). From the Hensen net toward four-dimensional biological oceanography. *Progress in Oceanography*, **56**: 7-136

Zarauz, L. (2007). Description and modelling of plankton biomass distribution in the Bay of Biscay by means of image analysis-based methods. *PhD Thesis.*: 164

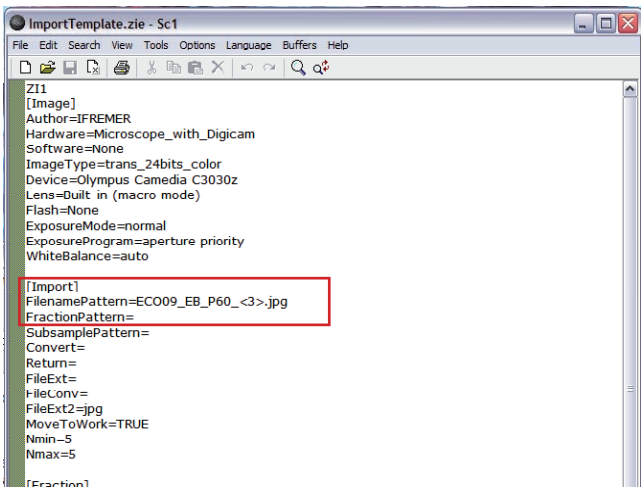
ANNEX I: Naming image files correctly (example)

In case of having one photo per sample (usually SCANNED IMAGES):

- Images: *ECO09_EB_P60_001.jpg...*

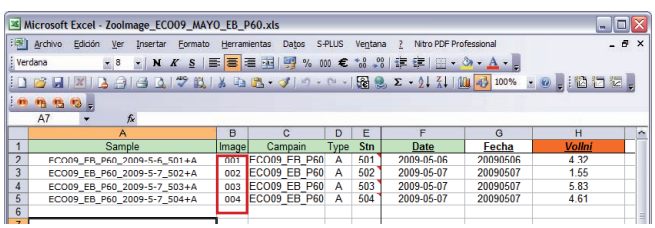


- ImportTemplate.zie: *"FilenamePattern": ECO09_EB_P60_<3>.jpg*



- In Excel file... (.txt file later on)
"Sample" column:
ECO09_EB_P60_2009-5-6_501+A
ECO09_EB_P60_2009-5-6_502+A
...

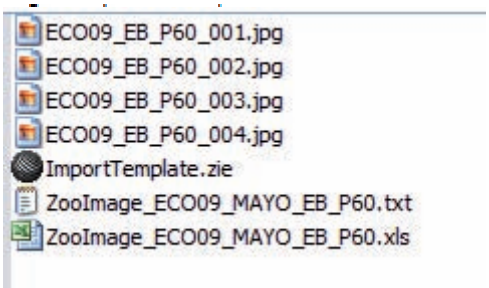
"Image" column:
001
002
...



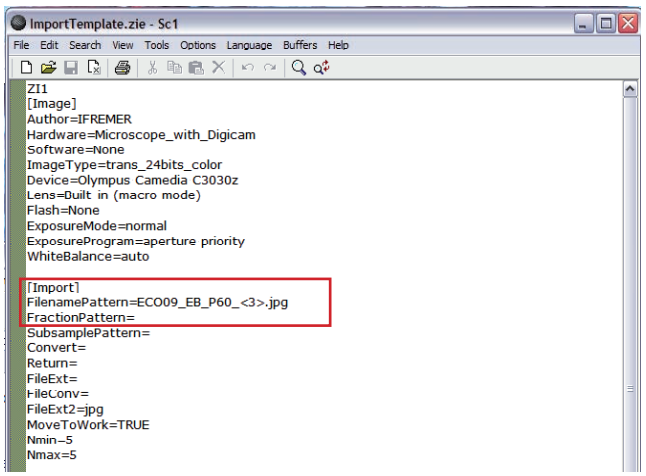
In case of having more than one photo per sample:

- Images: *ECO09_EB_P60_001.jpg...*

NOTE: All images are named equally, hence, it is essential to note which photo numbers correspond to each station, in order to define that in the *Excel* file for later automatic renaming.



- ImportTemplate.zie: *"FilenamePattern": ECO09_EB_P60_<3>.jpg*



- In Excel file... (.txt file later on)
"Sample" column:
ECO09_EB_P60_2009-5-6_501+A
ECO09_EB_P60_2009-5-6_502+A
...

"Image" column:
001-078
079-156

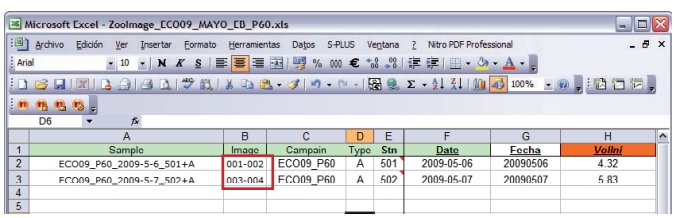
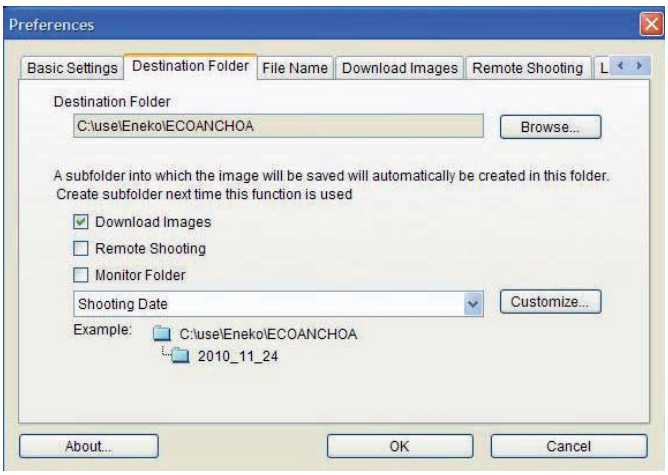


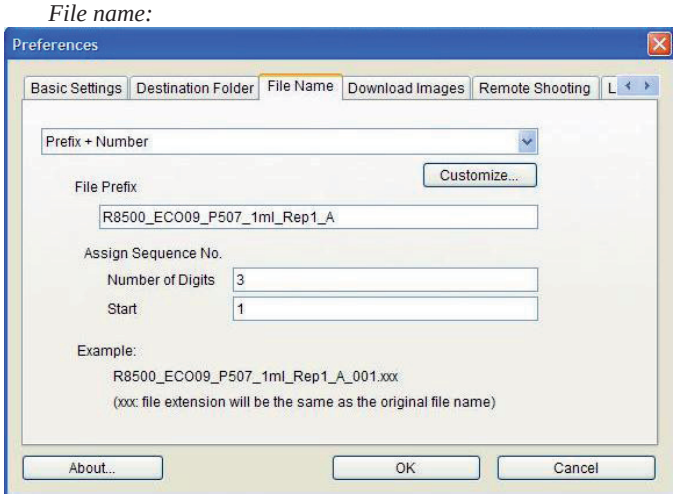
IMAGE NAMES with *EOS Utility*

Click on "*Preferences*" of the *EOS Utility* window to make changes both for file names or destination folder.

Destination folder: Where is required to place all images



File name:



NOTE: Final *.zid* files will not have "+A" at the end of the name. (e.g. *ECO09_EB_P60_2009-5-6_501.zid***)**

ANNEX II: Parameter definition for ZIM files

When digitizing images with *Zooimage*, **.zim files** will be created at first step, i.e. **one *zim* file per sampling station**. When opening one of those files, several parameters are defined, which should be checked sometimes in order to obtain correct labelled results. In this section, some of parameters commented are explained:

Vol.Ini : Initial volume (m^3)

It defines the volume of water filtered by each of the PAIROVET nets aboard the vessel.

NOTE: Vol.Ini should have at least **3 digits** of decimal precision.

SubPart: Aliquot volume (mL) / Volume of sample (mL)

It is the ratio between the aliquot and the initial sample volume (bottle of zooplankton, usually of 250 mL).

NOTE: Subpart should have at least **4 digits** of decimal precision.

PixelSize: Size (mm) of one pixel (i.e. measurement of **real resolution**)

It indicates the size of one pixel in millimetres. Each station or photo-group will have one defined **real resolution**. Checking sometimes this real resolution could be advisable.

- **In case of scanner**, pixel size could be calculated from this equation, since the resolution is previously known (defined by the scanner):
  Since 1 inch = 25.4 mm, the pixel size could be defined as:
  Pixel size = 25.4 / Resolution (dpi)
- **In case of photographs taken with camera**, this pixel size has to be measured with the *ImageJ* measuring the microrule size in the **calibration photo**. One *calibration photo* of the microrule was taken in each resolution.

NOTE: Put the microrule inside the plate, with the rule on the bottom side (next to the bottom of the plate)

- Open the *calibration photo* with *ImageJ*.
- Draw a line marking the millimetre of the microrule in the photograph.
- *Analyze → Measure*

Our marked line **length** will appear measured in **pixels**.

- For example, *with **900 dpi** configuration*: 38.13 pixel mm^{-1}
  Hence...
  1 pixel size = 1 / 38.13 **= 0.02622 mm**
- For example, *With **8500 dpi** configuration*: 348.1022 pixel mm^{-1}
  Hence...
  1 pixel size = 1 / 348.1022 **= 0.0028727 mm**

- Since the **resolution** is defined by the *amount of pixels in 1 inch (i.e. 25.4 mm),* our **real resolution** can be now easily calculated...
  - *With **900 dpi** configuration*:
    Resolution = 25.4 / 0.02622 **= 968.72 dpi**
  - *With **8500dpi** configuration*: 348.1022 pixel mm^{-1}
    Resolution = 25.4 / 0.0028727 **= 8841.85 dpi**

CellPart: Part of the sample area photographed (percentage)

It defined the part of the plate that is digitized. In our case, polystyrene plates have been measured at: 126 x 84 mm

Since digitized areas should not include borders, plate **inside area** is defined as: **10168 mm^2**

Hence...

- **In case of scanner**, *CellPart* could be calculated this way:
  - Open *ImageJ*
  - The size of the image appears at the bottom (Example: *11064 x 7500 pixels*)
  - Pixel size in case of scanned photos: Pixel size = 25.4 / Resolution (dpi)
  - As an example, for *2400 dpi resolution*:
    - Pixel size = 25.4 / 2400 = 0.0105833 mm
    - 11064 pixels = 117.094 mm
    - 7500 pixels = 79.375 mm
    - Photo area (mm^2) = 117.094 x 79.375 = 9294 mm^2
    - Hence the percentage of the plate represented in the photo (i.e. CellPart) would be:
      (9294x100) / 10168 = 91.4% → **CellPart = 0.914**
- **In case of photos taken with the digital camera**, *CellPart* could be calculated this way:
  - Open *ImageJ*
  - Open the microrule calibration photo (on plate) for this resolution
  - The size of the image appears at the bottom (Example: *11064 x 7500 pixels*)
  - The Pixel size can be calculated as it is explained in previous section
  - As an example, for *8500 dpi resolution*:
    - Pixel size=0.0028727 mm
    - 4272 pixels = 12.27 mm
    - 2848 pixels = 8.18 mm
    - Photo area (mm^2) = 12.27 x 8.18 = 100.368 mm^2
    - Hence the percentage of the plate represented in the photo (i.e. CellPart) would be:
      (100.368x100) / 10168 = 0.98%
      CellPart (of a single photo) = 0.0098

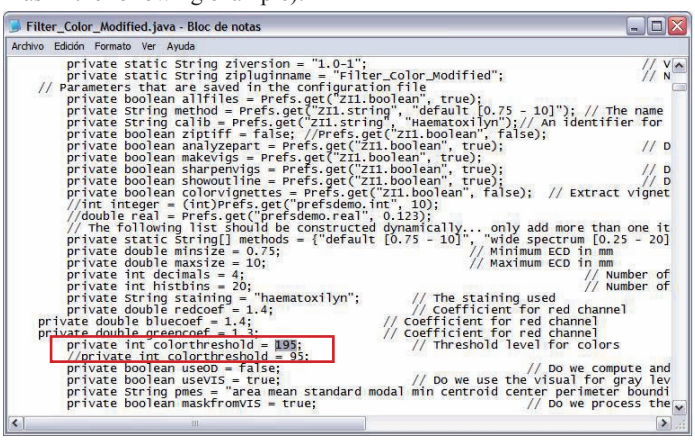
NOTE: In case of having more than one image for one station, in the corresponding ZIM file (one for all images of this sample or aliquot) the CellPart will be the whole value.

- Example:
  If 78 photos have been taken, of 100.368 mm^2 each, the CellPart value in the ZIM file would be:
  0.98% x 78 = 76.99
  CellPart (for .zim file) = 0.77

ANNEX III: Filter and threshold definition for image analysis

Filters used by *ImageJ* when extracting vignettes from digitized images are use to be compiled in a predefined way when using *Zooimage*. However, the way of working of those filters can be observed changing some parameters (i.e. filter code in *R* and/ or colour threshold values for each filter), making them more or less sensible when defining the silhouette of particles. Changing this *sensibility*, too many non-biological artefacts extracted as vignettes could be avoided, making later Training Set easier for the expert at the same time.

Opening a filter file with a text editor, the threshold number can be changed (changing the number of "*private int colorthreshold"* as in the following example):



NOTE: It is essential to copy selected filters into corresponding folder for further analyses: C:\Archivos de Programa\Zooimage\ bin\ImageJ\Plugins\Zoophytoimage\

Filter compilation

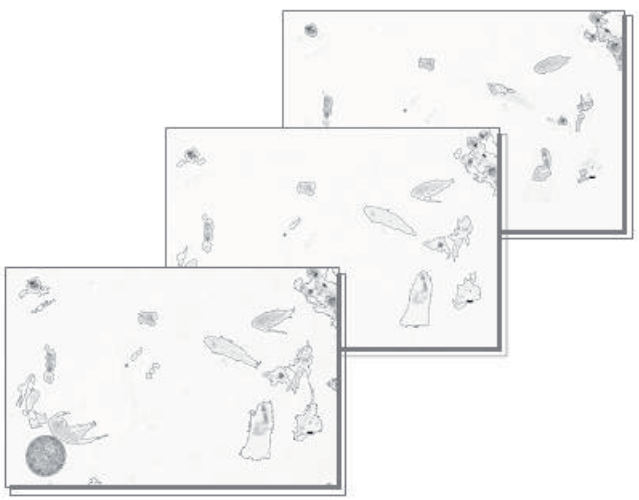
In case of Ichtyoplankton lab computer, there are **two *ImageJ* versions**:

- ***ImageJ***: Big images can be opened, but it does not compile new filters.
- ***ImageJCompile***: New filters can be compiled, but it does not allow opening big images.

To compile a different filter (using *ImageJCompile*):

- Open *ImageJCompile*
- *Plugins → Compile and Run*
- Choose the filter to compile (.java) → *Open*
- Now *ImageJ* asks for one image. Select **one .zim** file to open.

  NOTE: By clicking the option of "compile all samples of the folder (all .zim-s)", one new folder will be created for each station, with all created vignettes inside.
- Doing this, one image has been processed (or all images of corresponding station or .zim file), obtaining...
  - *RAW* folder: processed images (original images) and original .zim files
  - *WORK* folder: here many files can be found...
    - **_dat1.zim**
      This is the same .zim file that has been explained before, but in this case it **also** presents **particle size measurements**.
      NOTE: The particle amount could be seen in this file (total amount of particles counted).
    - **.out1** and **.vis1**
      These are auxiliary images, i.e. files extracted during the image analysis process. Open the *vis1.gif* file with *Paint*, and then paste the *out1.gif* as transparency above it, just to see which particles *ZooImage* has been taken into account during the process. In addition, the contour or silhouette assigned to each particle could be observed this way.
  - *ECO09_P508_1ml_Rep2_2400dpi* **folder**
    This is the folder where all extracted vignettes are collected at first.



Colour thresholds

Each colour threshold of the filter can be easily changed in order to compare differences in terms of total particle amount at the end of the process.

- Open an image with *ImageJ*
- *Image → Type → 8bit*
- *Image → Adjust → Threshold*
  The first value (up) is used to be 0 (or 10)

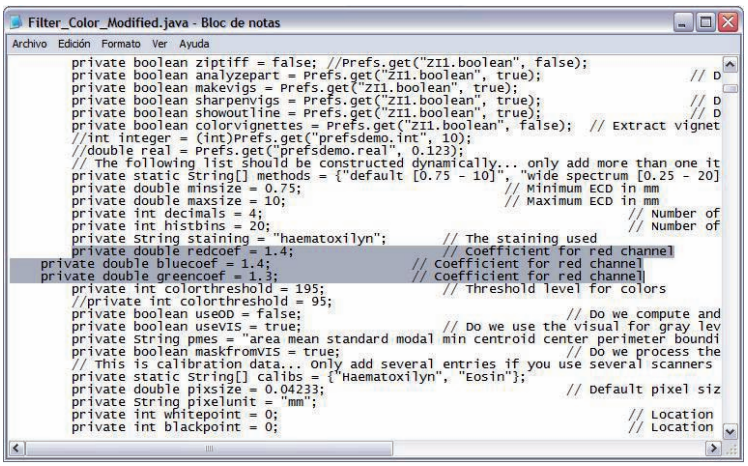
The second value (bottom) is the:
[number of threshold of filter file *.java*] + 50.
The image has three components that can be separated:

- Blue coef.
- Red coef.
- Green coef.

Those values could be modified in order to change the threshold for the image analysis. In case of **scanner** filter, there is no difference in those coefficients from one resolution to other. On the other hand, at the beginning of this work it is unknown if those values should be changed for **camera** photos...

Threshold value ↓ Amount of automatically counted particles ↓
Threshold value ↓ Area of each counted particle ↓

Opening a filter file with a text editor, colour threshold numbers can be also changed (changing the numbers of "*private double red/blue/green-coef*" as in the following example):



Total particles counted under microscope in case of 11 different samples were compared to total number of particles extracted with image analysis using different filter thresholds, in order to define the best levelled graph and so that best threshold (next page).

According to results obtained in different comparisons, **filters considered as optimum for each resolution and methods have been defined as:**

How does the colour filter work?

The colour image has at same time, three image-components (or colour component values):

- R (red component)
- G (green component)
- B (blue component)
- White photo: R=256 G=256 B=256
- Black photo: R=0 G=0 B=0

("black" means no color)

To make what filters use to do but in this case manually, those values can be also changed, observing this way the result obtained (image) in addition to threshold value changing.

- Open *ImageJ*
- *Image → Color → RGBsplit*

Now the three components of the image (in black & white) have been separated.

F.ex:

- RedCoef: 1.4
- GreenCoef: 1.3
- BlueCoef: 1.4

- *Process → Math → Multiply →* [the number of each component]
- *Image → Color → RGBmerge*

The three components have been now merged into one photo again (as the filter use to do itself).

- *Image → Type → 8bit*
- *Image → Adjust → Threshold*

Now, the threshold value (black & white) can be changed again.

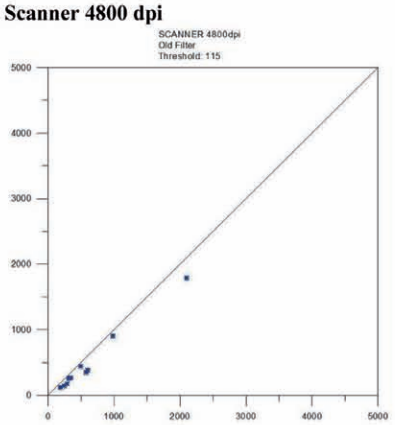
Imaging Device	Resolution (dpi)	Filter used	Colour threshold	Lab PC URL
SCANNER	4800	Scanner4800_Colour	115	*
CAMERA	4800	Camera4800_Colour	130	*
CAMERA	8500	Camera8500_Colour	110	*

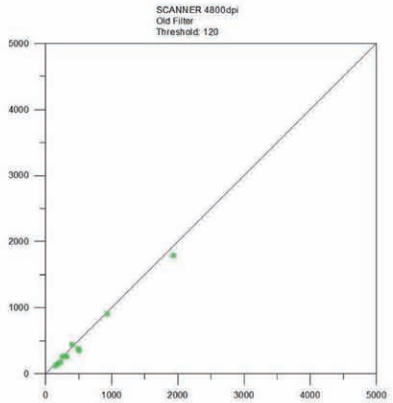
* C:\Program Files (x86)\ZooImage2\bin\ImageJ\plugins\ZooPhytoImage

NOTE: In case of the camera, a new filter has been predefined after threshold defining experiment: *Camera_Colour8500.java*: this is the one which has to be used from now in image analysis of samples digitized with the camera at the highest resolution (i.e. 8500 dpi).

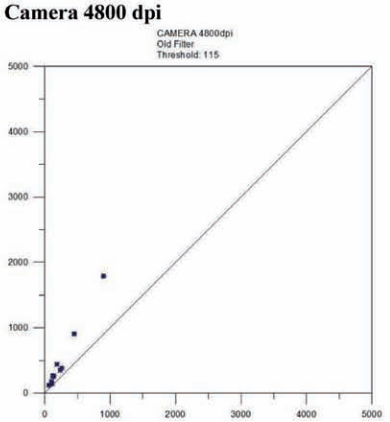
These filters and thresholds are predefined in lab computer systems in order to avoid confusion and any operator influence.

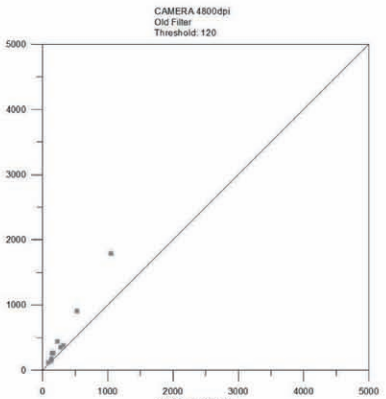
Scanner 4800 dpi

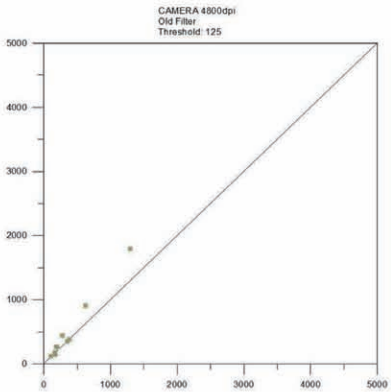


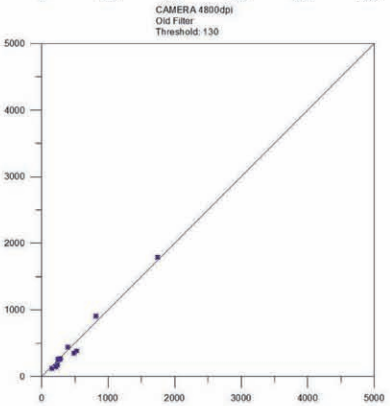


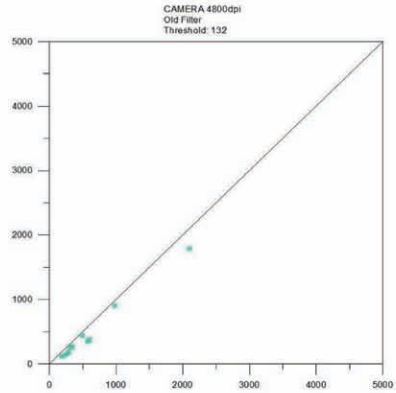
Camera 4800 dpi



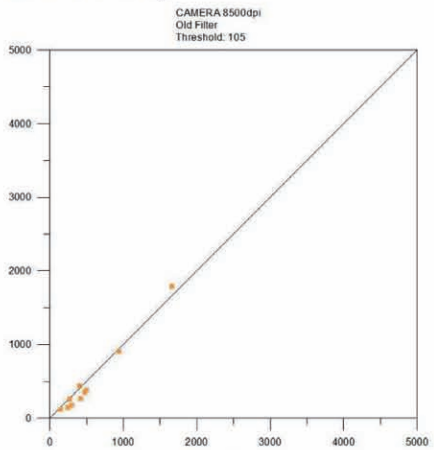


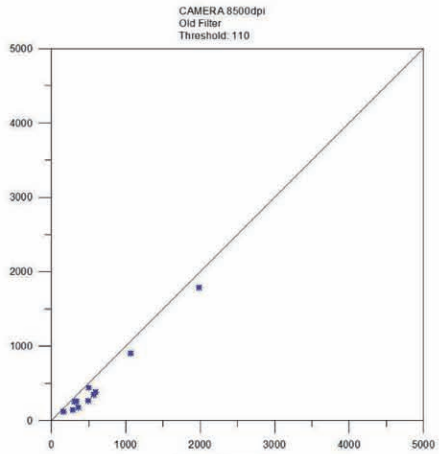


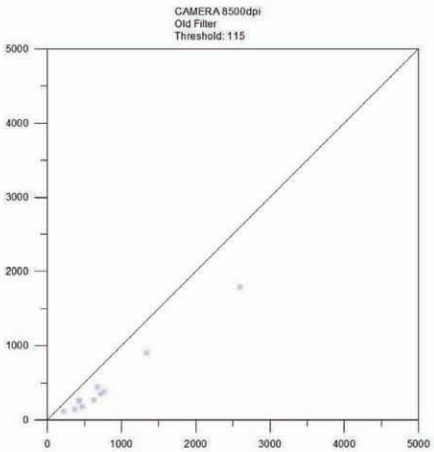


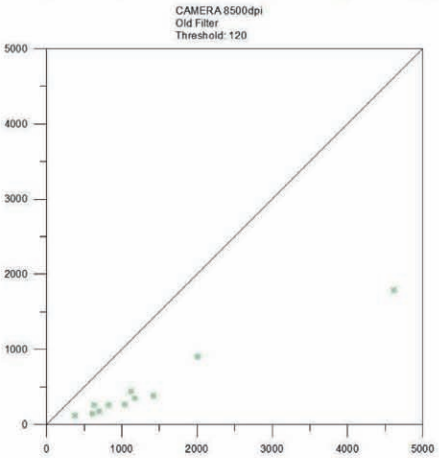


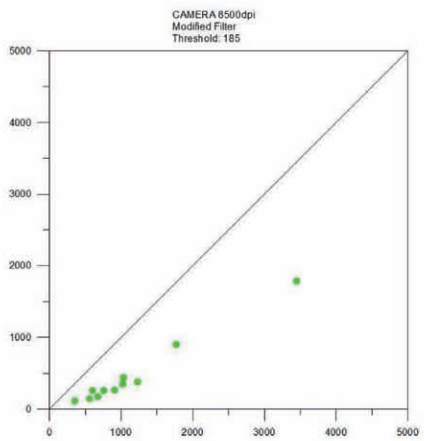
Camera 8500 dpi











ANNEX IV: Working with the Classifier

Interpretation of a Classifier analysis. Example.

	predicted				
classes	01	02	03	04	05
01 Acartia spp	76	0	0	1	0
02 ANE egg	0	8	0	0	0
03 Appendicularia Fritillaria sp	2	0	17	10	0
04 Appendicularia Oikopleura spp	0	0	5	86	0
05 Bivalve veliger	0	0	0	0	77
06 Calanoid small	17	0	0	0	0

CORRECT CLASSIFICATIONS:

- 76 *Acartia spp* have been classified correctly as *Acartia spp.*
- 8 *ANE eggs* have been classified correctly as *ANE eggs*.
- 17 *Appendicularia Fritillaria sp* have been classified as *Fritillaria sp.*
- 86 *Appendicularia Oikopleura spp* have been classified as *Oikopleura spp.*
- 77 *Bivalve veligers* have been classified as *Bivalve veligers*.

INCORRECT CLASSIFICATIONS:

- 1 *Acartia spp* has been classified as *Appendicularia Oikopleura spp.*
- 10 *Appendicularia Fritillaria sp* have been classified as *Oikopleura spp.*
- 2 *Appendicularia Fritillaria sp* have been classified as *Acartia sp.*
- 5 *Appendicularia Oikopleura spp* have been classified as *Fritillaria sp.*
- 17 *Calanoid small* have been classified as *Acartia sp.*

In this example, final abundances would be:

Acartia sp:

Correctly classified: 76
False negatives: -1 (classified as *Oikopleura spp*)
False positives: +2 (*Fritillaria sp*) +17 (*Calanoid small*)
ESTIMATED ABUNDANCE = 76-1+2+17 = 94

Appendicularia Oikopleura spp:

Correctly classified: 86
False negatives: -5 (classified as *Fritillaria sp*)
False positives: +1 (*Acartia spp*) +10 (*Fritillaria sp*)
ABUNDANCE = 86-5+1+10 = 92

Improving the Classifier

Many experiments have been done in order to improve the accuracy of the training set and so that the classifier.

METHOD 1: Training set of previously manually identified items (manual identification with a stereo microscope).

- At least 50 individuals per each species have to be caught in different recipients.
- Digitize all samples of manually pre-classified species.
- *.zim* files can be copied from previous analysis folder, but remember to change the name since it has to be the same as the image!
- Process this samples until create .**zid** files following instructions on this manual.
- Create a training set only with vignettes from manually identified images.
- Read the training set and save it as an object in *R.*

METHOD 2: Balance-imbalance problem

The amount of particles in each group of the training set has to be taken into account. In fact, criteria to define the maximum number of individuals per each group will be based on that can be found in real world. This way, common groups should have maximum defined amount of vignettes, whereas rare groups should have few particles. However, some common groups can show fragmented vignettes or they can appear as aggregates and that could suppose problems when finding vignettes of a defined class during the training set making.

- Maximum number of vignettes for abundant species will be defined once vignette extraction has been done and at first sight by the expert.
- All extra particles (exceeding predefined maximum number) will be placed in duplicated folders within "_" folder, for possible future vignette adding.
- In case of rare groups or taxonomic classes that have too few vignettes, **duplication** of same vignettes can be done, but it is not so recommended.
- Vignettes of **other training sets** can be also added in corresponding folder in order to increase the amount of vignettes in one class. In this case it is essential to **copy *dat1_.RData* files** from the source of adding vignettes (they are located inside *.zid* folders or in the main folder of the training set), otherwise those vignettes will not be taken into account.
- Vignettes should also be added in those classes that show high confusion level in the classifier (check the *Confusion Matrix* for that).
- Read the training set and save it as an object in *R.*

METHOD 3: Additional Artefact training set

- Duplicate the training set folder with all subdirectories and _ *dat1.RData* files and name it as "*_trainArtefacts*".
- Place all biological groups (previously classified) inside "_" folder and supervise those non-biological groups.
- Add artefact vignettes from "_" folder according to the *Balance-Imbalance* problem explained in previous section.
- Read the training set and save it as an object in *R.*

METHOD 4: Supervised automated classification of vignettes

Vignettes classified automatically by the classifier can be supervised manually by the expert. This way, some selected

vignettes (some can be very representative of a group, and can be classified in a wrong way) can be manually moved to corresponding correct place in our training set.

- It is necessary to charge ***ZIRes.r*** file in *R* before extracting results.
- When samples are processed to the last step, *_Automated* folder will be created in our main folder, next to our original training set folder ("*_train*").
- All *_dat1.RData* can be copied into the original training set folder. Once this has been made, vignettes (correctly classified of not) can be added from *_Automated* to the training set corresponding folder. This way vignettes correctly classified will improve their corresponding class, and wrong classified vignettes now located in their correct taxonomic group will be also considered as what they really are.
- Read the training set and save it as an object in *R*.

MERGING DIFFERENT TRAINING SETS

Once different training sets (i.e. improving training sets) have been made, both sets have to be merged, i.e. the original and the second one, which is going to improve it.

NOTE: It is essential to have **each training set saved** as an object in *R*.

- Open *ZooImage (R console)*
- Drag training set objects (i.e. original and additional) into *R Console* window
- Type the following code:

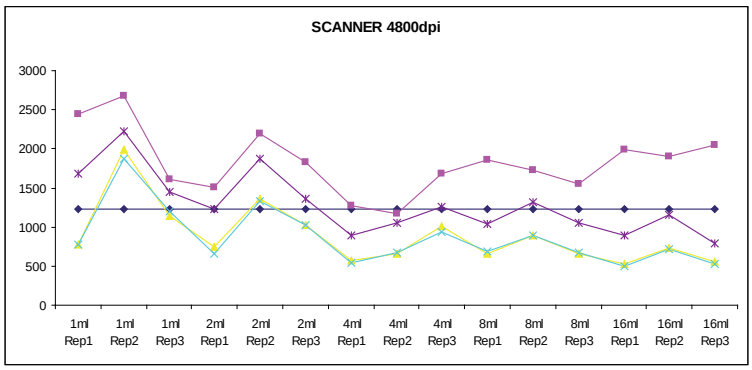
  ZItrain<-rbind(ZItrain1,ZItrain2)

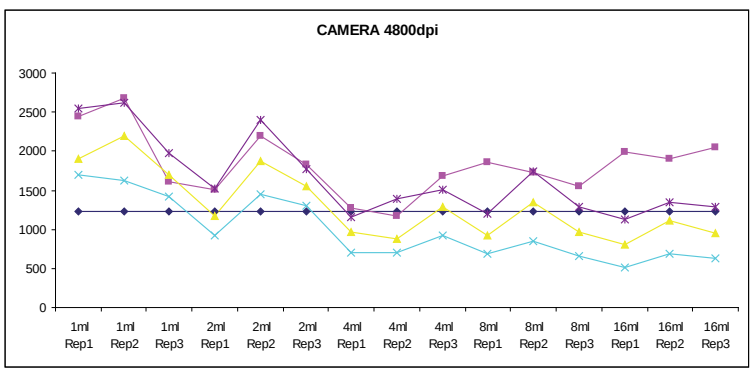
  Where...
  - *ZItrain* is the new object created from the merging of the two training sets,
  - *ZItrain1* is the original training set (f.ex.:"*_train*"),
  - *ZItrain2* is the additional training set made for improvement (f.ex.:"*_trainArtefacts*")
- Save the merged training set as an object in *R*.

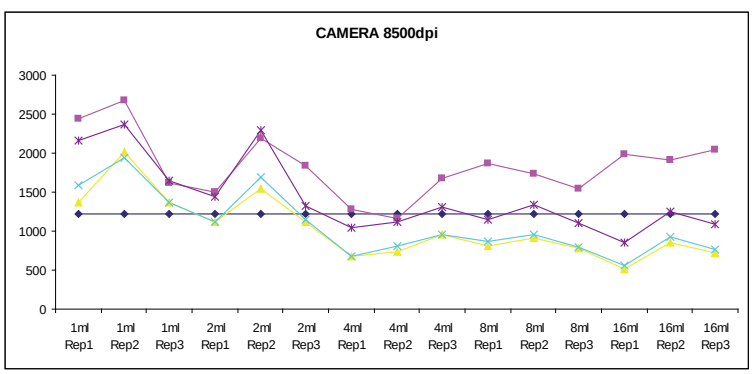
EXAMPLE: SOME RESULTS OF TRAINING SET IMPROVEMETS USING INTERNAL CONTROL METHODOLOGY FOR EVALUATION

Internal *Amberlite Bead* control methodology was applied in this experiment, in order to evaluate different training set improvements for automated analysis of images coming from different imaging devices

- Total particle (*Amberlite Bead*) number is represented as **abundance of beads m^{-3}.**
- *Theoretical* abundance does not change since different aliquots of the same sample have been taken for this experiment.
- *Manual Counting* abundances come from identification under stereo microscope. This is supposed to be the most real result so that differences from the theoretically predicted values could be due to manipulation or sampling process.
- *TS_Original*: Original training set, with artefact training improvement added. Maximum number of vignettes per group has been defined.
- *TS_ExpertCorrected*: Cleaning of classes has been done, selecting clearest vignettes and removing those that can cause confusion. Balance-imbalance correction has been also applied.
- *TS_ExpertCorrectedAfterClassification*: Supervised automated classification has been applied (Method 4 in this annex section) to *TS_ExpertCorrected.*







www.azti.es

azti tecnalia

Txatxarramendi ugartea z/g
48395 Sukarrieta (Bizkaia)
Tel.: +34 94 657 40 00
Fax: +34 94 657 25 55

Herrera Kaia, Portualdea z/g
20110 Pasaia (Gipuzkoa)
Tel.: +34 94 657 40 00
Fax: +34 94 657 25 55

Parque Tecnológico de Bizkaia
Astondo bidea. Edificio 609.
48160 Derio (Bizkaia)
Tel.: +34 94 657 40 00
Fax: +34 94 657 25 55