JOURNAL OF PLANKTON RESEARCH | VOLUME 31 | NUMBER 12 | PAGES 1505–1516 | 2009

Comparison between automated analysis of zooplankton using ZooImage and traditional methodology

ASTTHOR GISLASON* AND TERESA SILVA
MARINE RESEARCH INSTITUTE, SKULAGATA 4, PO BOX 1390, 121 REYKJAVIK, ICELAND

*CORRESPONDING AUTHOR: astthor@hafro.is

Received April 7, 2009; accepted in principle June 16, 2009; accepted for publication September 15, 2009; published online October 14, 2009

Corresponding editor: Mark J. Gibbons

The objective of this study was to evaluate the accuracy of the ZooImage image analysis system for taxonomic classification of zooplankton samples. For this purpose, automated analysis with the ZooImage software was compared with traditional analysis, using zooplankton samples collected in the Iceland Sea in July 2006. When compared with the traditional methodology, ZooImage was able to classify zooplankton into main taxonomic entities (size classes and families or genera in some cases), while being less successful in identifying the zooplankton into species. Other important information, that is difficult and time consuming to obtain by traditional methods such as biomass and size distributions are, however, easily obtained with ZooImage. The automated analysis takes much less time than the traditional methods. While the study confirms that ZooImage is a promising tool for rapidly analysing zooplankton samples, it is also clear that the traditional approach will be needed in future investigations, particularly studies addressing zooplankton community structure, development and life history.

INTRODUCTION

Zooplankton have a key role in marine ecosystems in linking primary production to higher trophic levels such as adult pelagic fish, and larvae and juveniles of pelagic and demersal fish. Their abundance and composition have therefore been widely studied since the earliest days of biological oceanography. Traditionally, zooplankton are taxonomically analysed by microscopic analysis of formalin preserved samples. However, this process is extremely time consuming and hence costly. In many cases, this severely limits the number of samples that can be processed, which in turn may lead to fragmented information that may be difficult to understand (Grosjean *et al.*, 2004).

Against this background, it is not surprising that significant efforts have been made to automate the analysis (e.g. Irigoien *et al.*, 2005; Benfield *et al.*, 2007). The early 1970s saw the advent of silhouette photography, as a way to produce permanent electronic records of zooplankton samples and to facilitate measurements and enumeration of individuals (Ortner *et al.*, 1979). However, this approach is not automatic as it involves manual processing of the photographic images, selecting and classifying of individuals for enumeration and counting. In the mid 1980s, image analysis systems were developed that were able to automatically count and measure particles, while however still not being able to identify the plankton to groups (Rolke and Lenz, 1984; Estep *et al.*, 1986). The early 1990s saw the advent of neural network algorithms for pattern recognition, which was a significant step forward, as they provided a way of teaching a computer to identify patterns (images) at high speeds and were relatively unaffected by imperfect images (Simpson *et al.*, 1992). The introduction of the random forest algorithm for machine learning in 2001 (Breiman, 2001) was a further improvement that should principally be the most efficient method for machine learning available at the time of its advent (Grosjean and Denis, 2007), with applications beyond zooplankton identification, for instance in medicine

doi:10.1093/plankt/fbp094, available online at www.plankt.oxfordjournals.org

(e.g. Shi *et al.*, 2005) and facial recognition (e.g. Kouzani *et al.*, 2007).

ZooImage is an open source software package using both the statistical software R and the image analysis software Image J for analysing images of zooplankton (http://www.sciviews.org/Zoo/PhytoImage) (Grosjean and Denis, 2007). Six different machine learning algorithms are imbedded in ZooImage (although more are implemented in R), among them neural networks and random forest. ZooImage is not limited to acquiring the images from one source only, and the images may be imported from relatively inexpensive devices like digital cameras or conventional office scanners. With the ZooImage package, one can calculate abundance, biomass and size spectra of zooplankton, classed by groups or species, and display the results as standardized per cubic meter of water filtered by the plankton net used for obtaining the samples (Grosjean and Denis, 2007). The approach takes significantly less time than the traditional methodology and thus, in addition to being a tool for analysing present day data sets, provides a promising tool for analysing historical plankton samples that otherwise would be hard to get analysed. A further advantage is that in regard to biomass estimation, the analysis is not destructive, as analyses for biomass tend to be, so the samples are kept after analysis (e.g. Alcaraz *et al.*, 2003; Zarauz *et al.*, 2007). An additional advantage is that the digital images are themselves permanent records of the physical samples that can be archived for storage on computers and even shared through the Internet.

To our knowledge, this is the third study to compare traditional and automated sample processing using ZooImage. While both the previous studies reported promising results, one of them (Plourde *et al.*, 2008) acknowledged that it was somewhat preliminary due to only six samples being included in the comparison, whereas the other (Bell and Hopcroft, 2008) did not use exactly the same zooplankton aliquots for the comparison. There are some further differences between the previous studies and the present one. Thus, both the earlier studies used the Epson Perfection 4990 Photo scanner for scanning the samples, while we use the newer model Epson Perfection V700 Photo Scanner. Given the somewhat different methodologies employed, a comparison of the results from these studies is interesting.

The objective of this study is to access the performance of ZooImage by comparing results of counting using ZooImage with those from microscopic analysis, using exactly the same aliquots for both approaches. The main aim is to evaluate the potential of the software as a tool for accessing the abundance and distribution of zooplankton by taxonomic groups. A further aim is to map the abundance of zooplankton in the Iceland Sea in July 2006.

METHOD

The sampling was undertaken in July 2006 at 17 stations in the Iceland Sea.

The zooplankton samples were collected using a HydroBios Multi Plankton Sampler (0.25 m^2 mouth area, 200 μm mesh size). The sampler was towed at a speed of ~1 ms^{-1}, usually from a depth of 300 m, collecting samples from five different depth layers. However, here we only deal with the samples collected in the uppermost layer (0–50 m). After collection, the zooplankton were transferred to glass jars and preserved in 4% formalin neutralized with sodium tetraborate.

Before the analysis, the zooplankton samples were subsampled with a Motoda splitter (Motoda, 1959), and an aliquot used for further analysis. The processing of the aliquots followed two paths, traditional taxonomic laboratory procedure using a stereomicroscope and automated analysis using the Epson perfection V700 Photo Scanner for image acquisition and the ZooImage software for analysing the images obtained.

Traditional analysis

The aliquots, usually containing ~100 *Calanus finmarchicus* and ~100 *C. hyperboreus*, the biomass dominant mesozooplankton species in this area, and at least 400 individuals of other species, were identified and counted under a stereomicroscope using the relevant identification literature.

Automated analysis

The same aliquots as analysed in the traditional manner were transferred to polystyrene trays (8.4 × 12.7 cm, ~107 cm^2) and water added to ensure that all animals were on the bottom of the plate and on the same plane level. Each aliquot usually had to be put into three to four trays. With a soft needle, the animals were then manually moved around the tray so as to ensure that as many animals as possible were inside a predefined cropping area of the plate. In this process, overlapping animals or animals touching each other were also separated manually from one another.

The aliquots were then scanned with the Epson Perfection V700 Photo Scanner with a dual lens system, using the Epson scan software to produce in an uncompressed format (tiff) 16 bit grey level images at 2400 dpi resolution (Fig. 1). The scanner was calibrated beforehand with respect to pixel size and range of grey level, following guidance given in the ZooImage Manual (Grosjean and Denis, 2007).

When scanning the samples, clean trays without any scratches were used, because in our experience dust and

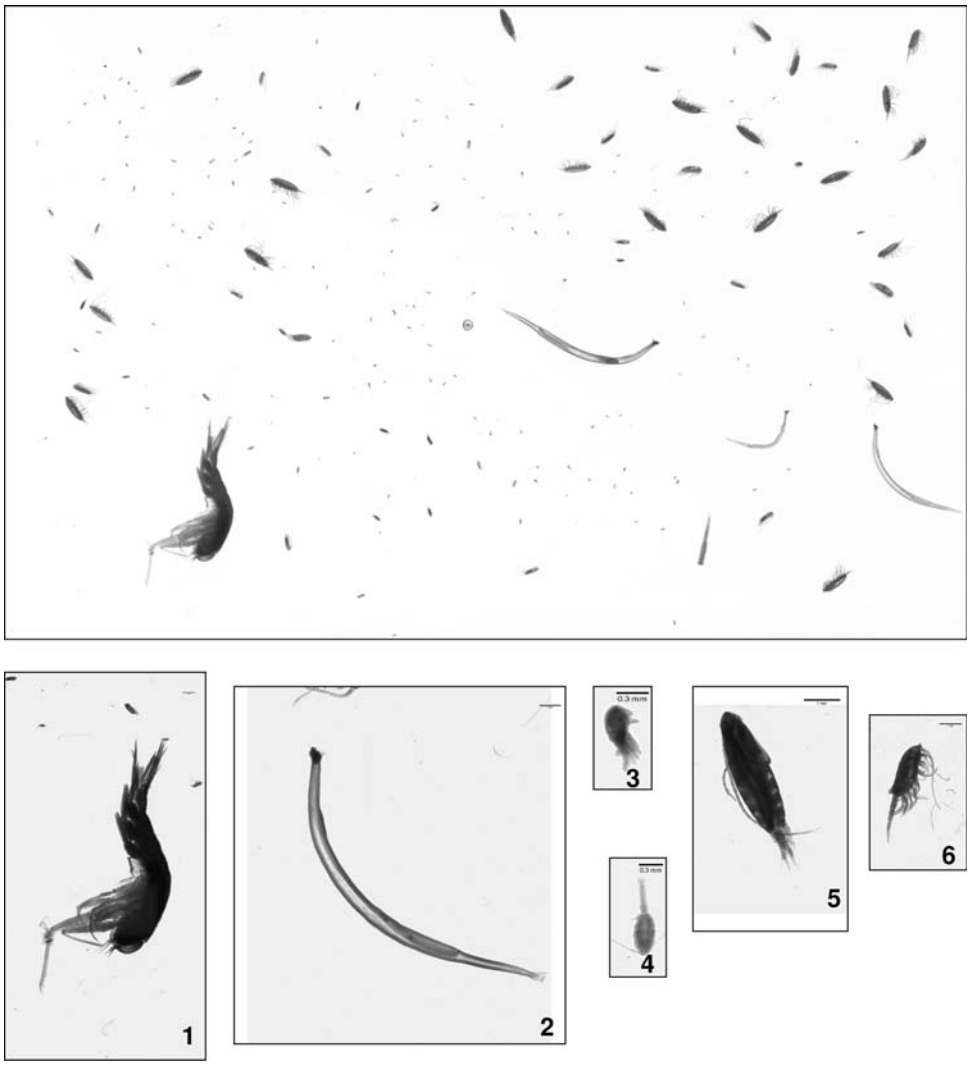

Fig. 1. Example of a scanned image (contents of one polystyrene tray after cropping, see main text for further details) produced by the Epson Perfection V700 Photo Scanner at 2400 dpi with 16-bit grey scale resolution, and a few vignettes extracted from it by ZooImage. (1, *Themisto libellula*; 2, chaetognatha; 3, *Oncaea* spp.; 4, *Oithona* spp.; 5, *Calanus finmarchicus* C6f; 6, *Metridia longa* C6f).

scratches will otherwise be interpreted as objects by ZooImage, thus creating an error. After scanning, the images were cropped using Adobe Photoshop Elements to remove the borders of the trays from the images. This inevitably led to a few animals being excluded from the final images. In order to correct for this, the animals outside the cropping area were counted from 17 images to arrive at an estimate of the average percentage of animals thus excluded by the cropping process ("cell part", Grosjean and Denis, 2007).

ZooImage version 1.2.1 (http://www.sciviews.org/zooimage/) was used to process all images using the plugin "Scanner Gray 16" to extract from a given image all the associated vignettes (i.e. small images of all the particles/animals containing a scale, Fig. 1). The plugin also extracts from each vignette 26 characteristics or features, related to grey level, size and shape, that are stored in a data file for each sample. This information is used by ZooImage in the machine learning process (see below), enabling the software to automatically estimate zooplankton number and biomass by taxonomic categories.

All the metadata concerning the image (cruise, station, location, date, sampled volume, fraction subsampled, calibration and so on) are stored in a metadata file by ZooImage.

The first step in training the computer to identify vignettes is to make a training set, where a representative subset of the vignettes is classified manually into taxon categories or groups. The training set and the six classification algorithms that are included in ZooImage are then used to build classifiers, which are used by the

computer to identify the vignettes. For this study, three different training sets were established consisting of 19–34 categories, with at least 15–50 vignettes in each (Table I). When building the training sets, care was taken to ensure that all the different objects (vignettes) were assigned to a category or group, even bubbles, fibers, detritus, marine snow, parts of organisms and vignettes with more than one particle/organism. Further, an attempt was made to include a variety of vignettes in each group (i.e. copepods of different orientation were included in the copepod group), so as to "teach" the computer the variability inherent within a particular group.

At the end of the learning phase, the performance of the classifier was assessed by calculating a so-called 10-fold cross-validation matrix between the manual and the automated classifications (Grosjean and Denis, 2007). In this process, the original training set is randomly partitioned into 10 parts (folds), and the classifier is created and evaluated 10 times, each time using a different part for evaluation while the remaining 9 parts are put together to create the classifier. The results of the assessment for all categories in the training set, as well as the overall average accuracy by category, is illustrated by a 10-fold cross-validated confusion matrix between the manual and the automated classifications. The confusion matrix is a square contingency table that compares all groups of the manual classification with all groups of the automatic recognition. The average accuracy (% of correctly classified items) across all 10 trials is computed as an estimator of classifier performance.

A comparison between the automated and manual analyses was made by comparing distribution maps based on data derived from the two approaches and by carrying out linear regression analyses between log-transformed abundance estimates by the two methods. For this comparison, the classes that resulted in poor classification (accuracy $<75\%$) were omitted, and groups that were not sampled quantitatively by the 200 μm mesh WP2 net (Bivalvia, Phytoplankton, Gastropoda, Protozoa, unidentified items), as well as artifacts (air bubbles, fibers, scratches and shadows).

RESULTS

Training sets and classifier performance

For this study, three training sets were set up, resulting in three classifiers being made based on the random forest algorithm (Table I). Originally, a detailed training set was established consisting of 34 groups, with 10–121 vignettes in each, so as to reflect as much as possible the species composition of the samples (as revealed by the traditional microscopic analyses of the samples). Initial tests with different classification algorithms revealed; however, relatively low accuracy with respect to some of the calanoid copepods (e.g. *Pseudocalanus* spp. C6f, *Metridia* spp., *Calanus* spp. C1, C2 and C6f) (Table I), so a further training set was created where the calanoid copepods were classed by sizes (basic training set). This resulted in four size groups of calanoid copepods, all being detected with acceptable accuracy by the classifier (Table I). The basic training set had 25 groups with 16–208 vignettes in each. In setting up the third training set (simple training set), we combined those zooplankton groups that had fewer than 50 vignettes into one special group ("Other Zooplankton"), resulting in a training set of 19 groups with 17–212 vignettes in each (Table I).

Depending on which algorithm was used in creating the classifier from the training sets, large differences in the 10-fold cross-validation accuracy were observed (Fig. 2). Random forest was the best algorithm for all training sets, with neural network and linear discriminant analysis coming next and performing similarly well (Fig. 2). With respect to random forest, the basic training set produced similar accuracy as the simple training set, and as it also was able to resolve the samples in more detail (Table I), we decided to use it for further analysis of the data.

In order to investigate whether the number of vignettes in each category of the training set has an effect on classifier performance, least squares regression analyses were carried out for the three training sets classed by the random forest algorithm with the number of vignettes per category as independent variable and ZooImage accuracy as dependent variable (Fig. 3). Both linear and non-linear models (Ivlev functions) were attempted, the latter (i.e. an exponential rise to the maximum) probably being most appropriate as the accuracy cannot exceed 100%. Only for the detailed classifier did we find a significant relationship between the number of vignettes and ZooImage accuracy, where the Ivlev model explained 9.8% more of the variance than the linear regression (Fig. 3).

Table II shows the confusion matrix for the basic training set when the random forest algorithm was used to build the classifier. Each row of the matrix represents the manual identifications, whereas the columns represent the automatic ones. The numbers in the cells correspond to the counting of vignettes, and from them the success of the classifier in identifying the vignettes may be inferred. For instance, 1 gastropod was misidentified by the automatic classifier as a bivalve, 1 as air bubble, 2 as unidentified, whereas 19 were correctly

Table I: Composition of three training sets that were made to train the computer and build the corresponding classifiers using the random forest algorithm

Detailed			Basic			Simple		
Class ($n = 34$)	Number of vignettes	Accuracy	Class ($n = 25$)	Number of vignettes	Accuracy	Class ($n = 19$)	Number of vignettes	Accuracy
Protozoa	19	1.00	Protozoa	19	0.95	Calanoida <1 mm	212	0.92
Polychaeta	16	0.69	Polychaeta	16	0.69	Calanoida 1–2 mm	75	0.83
Gastropoda	23	0.87	Gastropoda	23	0.83	Calanoida 2–3 mm	88	0.92
Bivalvia	16	0.94	Bivalvia	16	0.94	Calanoida >3 mm	39	0.95
Calanus finmarchicus C1	31	0.58	Calanoida <1 mm	208	0.93	Copepoda nauplii	85	0.91
Calanus finmarchicus C2	22	0.64	Calanoida 1–2 mm	76	0.83	*Oithona* spp.	75	0.81
Calanus finmarchicus C3	35	0.83	Calanoida 2–3 mm	88	0.91	*Oncaea* spp.	53	0.85
Calanus finmarchicus C4	25	0.84	Calanoida >3 mm	39	0.92	Chaetognatha	58	0.95
Calanus finmarchicus C5	42	0.67	Copepoda nauplii	85	0.89	Larvacea	55	0.60
Calanus finmarchicus C6F	18	0.50	*Oithona* spp.	75	0.88	Other zooplankton	143	0.69
Calanus hyperboreus C1	11	0.27	*Oncaea* spp.	39	0.79	Unidentified	41	0.85
Calanus hyperboreus C2	10	0.20	Amphipoda	21	0.57	Marine snow	36	0.53
Calanus hyperboreus C3	25	0.84	Euphausiacea	29	0.24	Molts	40	0.53
Calanus hyperboreus C4	38	0.92	Chaetognatha	40	0.88	Phytoplankton	19	1.00
Copepoda nauplii	85	0.89	Larvacea	55	0.60			
Pseudocalanus spp. C1–C5	121	0.86	Ophiuroidea	19	0.63	Bubble	24	0.88
Pseudocalanus spp. C6F	16	0.13	Unidentified	41	0.90	Detritus	17	0.47
Metridia spp.	11	0.09	Marine snow	36	0.61	Fiber	42	0.81
Oithona spp.	75	0.84	Molts	40	0.53	Scratch	24	0.75
Oncaea spp.	41	0.80	Phytoplankton	19	1.00	Shadow	48	0.90
Amphipoda	21	0.57						
Euphausiacea	29	0.41	Bubble	24	1.00			
Chaetognatha	40	0.90	Detritus	17	0.41			
Larvacea	55	0.67	Fiber	42	0.81			
Ophiuroidea	19	0.63	Scratch	24	0.75			
Unidentified	41	0.90	Shadow	48	0.92			
Marine snow	36	0.56						
Molts	40	0.50						
Phytoplankton	19	1.00						
Bubble	24	1.00						
Detritus	17	0.41						
Fiber	42	0.81						
Scratch	24	0.75						
Shadow	48	0.92						
Total	1135			1139			1174	

Number of individuals in each class and accuracy as estimated by 10-fold cross-validation (% of correctly classified items across 10 trials) is also given.

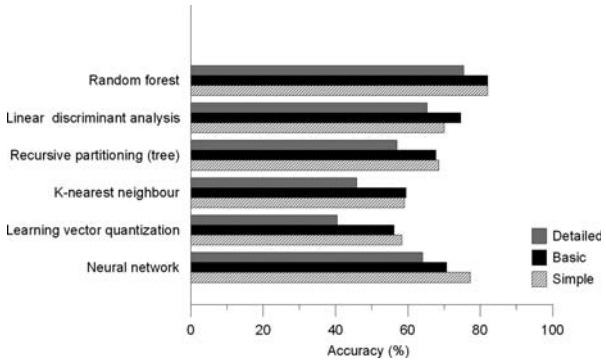

Fig. 2. The accuracy of classifiers based on three training sets (detailed, basic and simple) using six algorithms as estimated by 10-fold cross-validation (% of correctly classified items across 10 trials).

identified (Table II). The diagonal line (from top-left to bottom-right) thus represents the correct counting of predicted vignettes, while the cells outside of the diagonal depict errors made by the automatic classifier (provided there is no error in the manual training set).

From the confusion matrix, it is evident that the automated classifier is least successful in identifying amphipods, detritus, euphausiids, larvaceans, marine snow, molts, ophiuroideans and polychaetes (accuracy $<75\%$), whereas the other groups are identified with good accuracy ($\geq 75\%$) (Table II).

Comparison between manual and automated sample processing

Figure 4 compares patterns of zooplankton distribution in the Iceland Sea in July 2006 as estimated by manual processing and analysis with ZooImage using the classifier based on the basic training set and the random forest classification algorithm. From the figure, it is evident that ZooImage did a relatively good job in identifying all the groups, with the same general trends being shown by the two methodologies. However, for some groups, ZooImage appeared to overestimate slightly the abundance, whereas for others it underestimated it. Thus, ZooImage underestimated the abundance of the "Calanoida <1 mm" group, whereas for the other size groups of calanoid copepods including copepod nauplii, ZooImage had a tendency to overestimate the abundance (Fig. 4). With regard to the cyclopoid copepods *Oithona* spp. and *Oncaea* spp., ZooImage had a tendency to underestimate the former and overestimate the latter (Fig. 4). Chaetognaths were overestimated by ZooImage (Fig. 4).

In order further to compare the agreement between the two methods, Fig. 5 presents linear regressions between the manual and automatic counting using the same data set (log $x+1$ transformed) as shown in Fig. 4. A one-to-one relationship, with the regression line going through the origin, would imply that the two methods were in perfect harmony. An inspection of the coefficients of determination of the linear regression equations (Table III) confirms the generally good agreement between the two methodologies for analyzing the samples as described in the previous section.

DISCUSSION

Training sets and classifier performance

The establishment of a training set and the creation and evaluation of the resulting classifier are generally considered the most critical stage in the automated processing of samples with ZooImage (Grosjean *et al.*, 2004; Grosjean and Denis, 2007; Plourde *et al.*, 2008; Bell and Hopcroft, 2008). In the present study, we evaluated the performance of three classifiers, and concluded that the one with the medium number of classes and the medium number of vignettes on average in

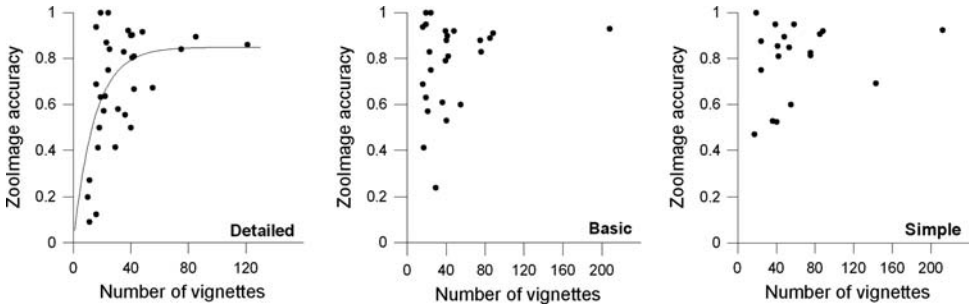

Fig. 3. Relation between the number of vignettes and ZooImage accuracy for detailed, basic and simple training sets, with 34, 25 and 19 categories, respectively, classed by the random forest algorithm. The line in the first panel (detailed) represents a non-linear least-squares Ivlev fit of ZooImage accuracy (Acc) to number of vignettes (Vig), $\mathrm{Acc}=0.849\times(1-\mathrm{E}^{(-0068^{*}\times\ \mathrm{Vig})}$, $R^2=0.32$, $P<0.001$.

Table II: Confusion matrix for the basic training set classed by the random forest algorithm

True Class \ Predicted	Amphipoda	Bivalvia	Bubble	Calanoida 1-2 mm	Calanoida 2- 3 mm	Calanoida >3 mm	Calanoida <1 mm	Chaetognatha	Copepoda nauplii	Detritus	Euphausiacea	Fiber	Gastropoda	Unidentified	Larvacea	Marine snow	Molt	*Oithona* spp.	*Oncaea* spp.	Ophiuroidea	Phytoplankton	Polychaeta	Protozoa	Scratch	Shadow	Total	P_D
Amphipoda	12	0	0	2	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0.57
Bivalvia	0	15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	16	0.94
Bubble	0	0	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	1.00
Calanoida 1-2 mm	0	0	0	63	9	0	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76	0.83
Calanoida 2- 3 mm	0	0	0	5	80	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	88	0.91
Calanoida >3 mm	0	0	0	0	3	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0.92
Calanoida <1 mm	0	0	0	0	0	0	194	0	5	0	0	0	0	0	4	0	0	5	0	0	0	0	0	0	0	208	0.93
Chaetognatha	1	0	0	2	0	0	0	35	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	40	0.88
Copepoda nauplii	0	0	0	0	0	0	3	0	76	0	0	0	0	0	1	0	1	4	1	0	0	0	0	0	0	85	0.89
Detritus	0	0	0	0	0	0	1	0	2	7	0	0	0	0	1	0	2	0	0	0	0	2	1	0	0	17	0.41
Euphausiacea	0	0	0	8	1	0	12	0	0	0	7	0	0	0	1	0	1	0	0	0	0	0	0	0	0	29	0.24
Fiber	0	0	0	0	0	0	0	2	0	0	0	34	0	0	3	0	3	0	0	0	0	0	0	0	0	42	0.81
Gastropoda	0	1	1	0	0	0	0	0	0	0	0	0	19	2	0	0	0	0	0	0	0	0	0	0	0	23	0.83
Unidentified	0	0	1	0	0	0	0	0	0	0	0	0	3	37	0	0	0	0	0	0	0	0	0	0	0	41	0.90
Larvacea	0	0	0	2	0	0	6	5	2	0	2	0	0	0	33	0	4	1	0	0	0	0	0	0	0	55	0.60
Marine snow	0	0	0	1	0	0	1	0	0	0	0	2	0	0	1	22	7	0	0	0	1	0	0	0	1	36	0.61
Molt	0	0	0	1	0	0	1	2	0	0	1	1	0	0	4	7	22	0	0	0	0	0	0	0	1	40	0.55
Oithona spp.	0	0	0	0	0	0	2	0	1	0	0	0	0	0	1	0	0	66	5	0	0	0	0	0	0	75	0.88
Oncaea spp.	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	5	31	0	0	0	0	0	0	39	0.79
Ophiuroidea	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	4	0	12	0	0	0	0	0	19	0.63
Phytoplankton	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	19	1.00
Polychaeta	0	0	0	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	1	0	0	11	0	0	0	16	0.69
Protozoa	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	18	0	0	19	0.95
Scratch	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	18	3	24	0.75
Shadow	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	0	0	0	0	0	0	0	44	48	0.92
Total	13	16	26	83	96	42	229	46	90	7	11	39	23	40	52	31	40	85	38	12	20	13	19	18	49	1139	
SP	0.92	0.94	0.92	0.76	0.83	0.86	0.85	0.76	0.84	1.00	0.64	0.87	0.83	0.93	0.63	0.71	0.55	0.78	0.82	1.00	0.95	0.85	0.95	1.00	0.90		

Each row of the matrix represents the groups present in the training set, whereas the columns show how these were classed by ZooImage. The diagonal line (from upper left to bottom right) shows the correct classification of vignettes. P_D (Probability of detection) is the probability that individuals will be correctly identified by ZooImage; SP (Specificity) is the probability that ZooImage predictions were correct for each taxon (terminology from Hu and Davis, 2006). The numbers in the cells indicate the number of vignettes, and the colors indicate the percentage of the total numbers of vignettes in each class: yellow (>0–10%), orange (10–20%) and red (20–100%).

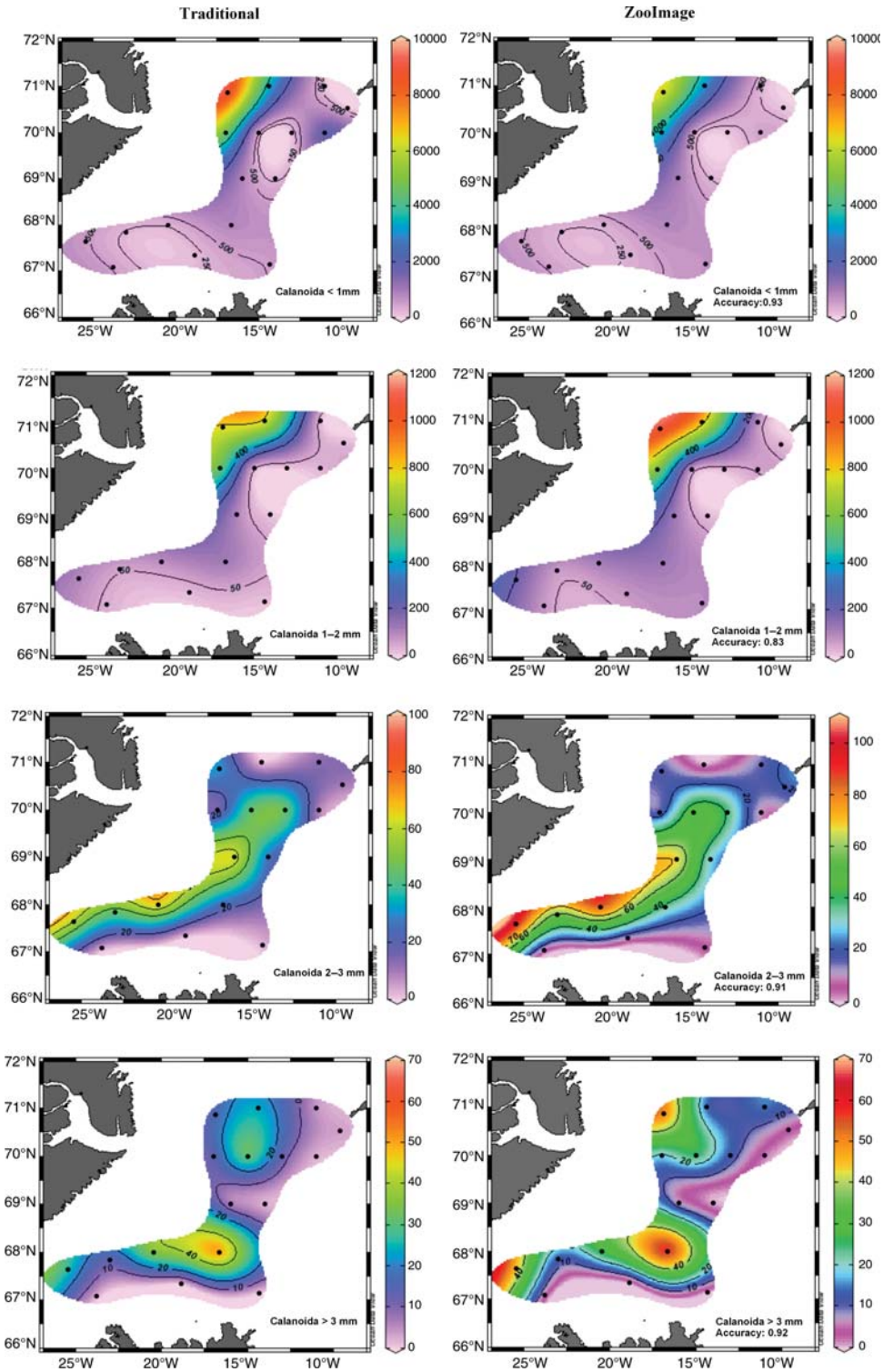

Fig. 4. Comparison of the abundance of four size groups of calanoid copepods (<1 mm, 1–2 mm, 2–3 mm, >3 mm), copepod nauplii, two cyclopoid groups (*Oithona* spp., *Oncaea* spp.) and chaetognaths (numbers m^{-3}, 0–50 m) in the Iceland Sea in July 2006 as estimated by traditional (left panel) and ZooImage (right panel) analysis. Note different scales on the panels for different groups.

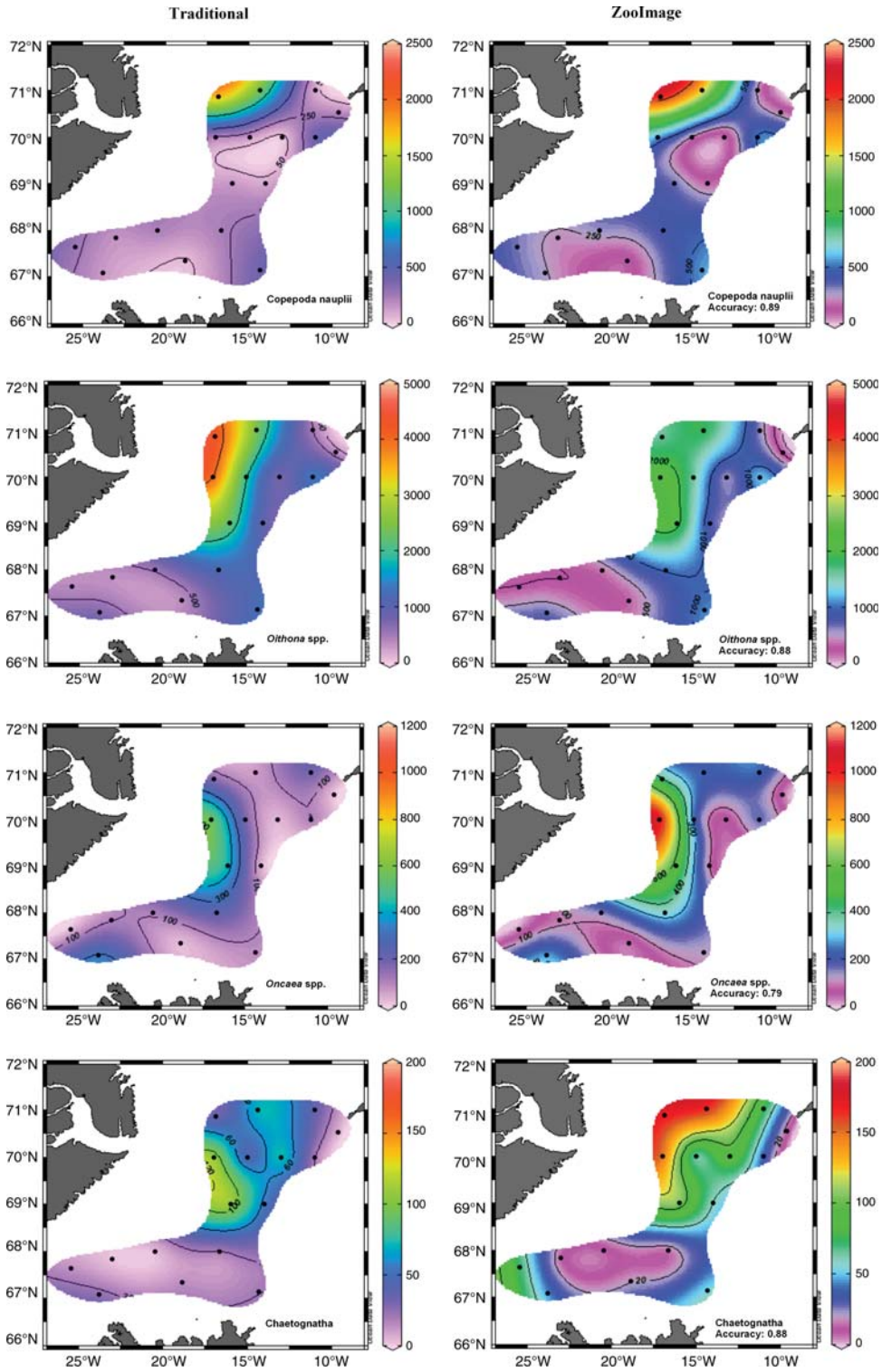

Fig. 4. Continued.

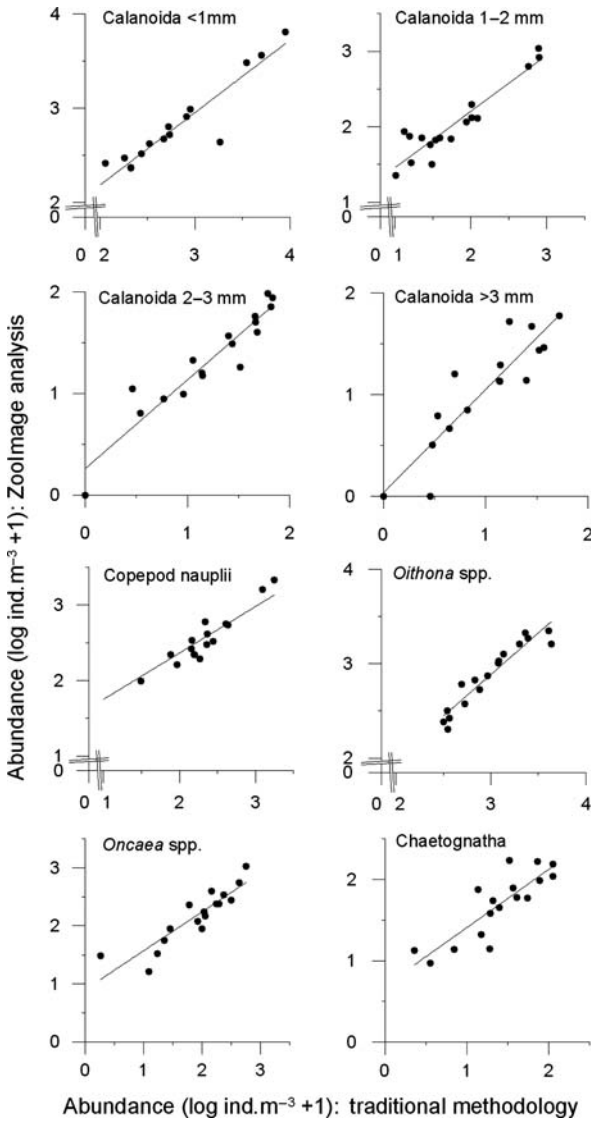

Fig. 5. Scatterplots of log-transformed abundance estimates based on manual classifications and automated ones using ZooImage for four size groups of calanoid copepods (<1 mm, 1–2 mm, 2–3 mm, >3 mm), copepod nauplii, two cyclopoid groups (*Oithona* spp., *Oncaea* spp.) and chaetognaths. The straight lines represent best fits of the linear regressions.

Table III: Results from linear regression analyses between abundance estimates (log ind. m^{-3} +1) using automated (ZooImage) (dependent variable) and manual (independent variable) analyses of the same aliquots

Group	Coefficient	*b*	*n*	*F*	R^2	*P*-value
Calanoida <1 mm	0.7688	0.6468	17	148.90	0.9085	<0.0001
Calanoida 1–2 mm	0.7484	0.7054	17	101.00	0.8707	<0.0001
Calanoida 2–3 mm	0.8745	0.2593	17	124.40	0.8924	<0.0001
Calanoida >3 mm	1.0157	0.0355	17	78.86	0.8402	<0.0001
Copepoda nauplii	0.6151	1.1333	17	87.68	0.8538	<0.0001
Oithona spp.	0.8816	0.2409	17	139.20	0.9027	<0.0001
Oncaea spp.	0.6705	0.8983	17	59.56	0.7988	<0.0001
Chaetognatha	0.7155	0.6930	17	37.25	0.7129	<0.0001

The regression coefficients (Coefficient), intercept (*b*), sample size (*n*), *F*-value (*F*) and proportion of variance explained (R^2) are given together with their respective significance levels (*P*).

each, generally performed the best (Table I; Fig. 2). Other similar studies have shown that the accuracy tends to increase as the number of classes decreases, while also being dependent on the nature and number of vignettes in the classes (Culverhouse *et al.*, 2003; Grosjean and Denis, 2007; Fernandes *et al.*, 2009). When determining the number of classes that a training set should contain, a compromise has to be made between the number of zooplankton groups that are to be identified automatically and the accuracy of the identifications. Other types of costs not considered here, such as costs and labor of establishing the training set, may even need to be taken into account. The decision to go for the basic classifier instead for the detailed one is at the cost of loosing taxonomic resolution, while keeping important information on size structure of the most abundant group, the copepods, which is a meaningful parameter in the ecological sense, for instance, with regard to trophic ecology.

The decision to establish the simple training set, by combining those classes that contained the fewest number of vignettes into one ("Other zooplankton") was taken as based on the experiences of Davis *et al.* (Davis *et al.*, 2004), who found that having classes with fewer than 50 images per category in the training set tended to decrease classifier performance. The simple training set had more than 50 images for most categories (Table I), but still the performance of the simple classifier was not any better than that of the basic one (Fig. 2). The reason for this apparent discrepancy between the present study and that of Davis *et al.* (Davis *et al.*, 2004) may be related to the fact that the two studies used different types of software for the machine learning process, with ZooImage being used in the present study, whereas Davis *et al.* (Davis *et al.*, 2004) used Visual Plankton (Davis *et al.*, 2004). The difference could also arise because image quality or the number of groups was not the same between the studies. It may be noted in this regard, however, that the number of vignettes that should be included per category of the training set may also depend on the number of categories (Fig. 3). According to Gashler *et al.* (Gashler *et al.*, 2008), random forest does poorly when faced with

irrelevant attributes of the items to be classed, and it may be argued that the number of vignettes with features that are irrelevant to a particular category is likely to be higher in the detailed training set (because the categories were so similar) than in the other two (simple and basic training sets). Increasing the number of vignettes in each category may thus have been particularly effective for the detailed training set in increasing the strength of the classifier (Fig. 3).

Random forest was the best classifier algorithm for all training sets. While there are a number of different classification algorithms and combinations of these available (see Grosjean *et al.*, 2004 for an overview), the random forest tends to score among the highest in accuracy (Grosjean *et al.*, 2004; Bell and Hopcroft, 2008; Fernandes *et al.*, 2009; Irigoien *et al.*, 2009).

From the confusion matrix, it is evident that the automated classifier was least successful in identifying amphipods, detritus, euphausiids, larvaceans, marine snow, molts, ophiuroideans and polychaetes (accuracy $<75\%$) (Tables I and II). The training set for several of these classes contained relatively few items each (16–55, Table I), and it is possible that the poor performance with respect to these groups was related to this. However, the poor performance may also be related to the characteristics of the vignettes contained in the training set for these poor performance classes. A number of morphological and image measurement features are extracted from the vignettes during the image analysis process and as pointed out by Fernandes *et al.* (Fernandes *et al.*, 2009), it may be difficult to establish how the different features are used by the classification algorithms during the machine learning process. It seems, however, evident that the shape of the items may not be the most important feature in this respect as an inspection of the confusion matrix reveals that misclassification occurs between animals of very different shapes and morphologies. Amphipods and euphausiids were, for instance, mostly confused with the very dissimilar group calanoid copepods, as were *Oithona* spp. and ophiuorideans (Table II). It is noteworthy in this respect, that the groups Calanoida <1 mm and Calanoida >3 mm, that generally have similar shapes were not confused by the classifier (Table II). This seems further to indicate that machine learning is a complex process simultaneously taking into account a number of properties of the images, not least size, as also pointed out by Bell and Hopcroft (Bell and Hopcroft, 2008).

When evaluating the accuracy with respect to the different classes, it is worth pointing out that the fact that the number of items in each class was different may have introduced bias in the sense that the different classes were estimated with different accuracy (Plourde *et al.*, 2008). A further point is that humans make mistakes (Culverhouse *et al.*, 2003), and the performance of the classifier would ultimately depend on how good the specialist is in identifying and labeling the vignettes when establishing the training set.

Comparison between manual and automated sample processing

Comparison between manual and automated processing of field samples show that ZooImage slightly underestimated the abundance of the Calanoida <1 mm group (Fig. 4), probably mainly because it misidentified some of them for *Oithona* spp. or even copepod nauplii (Table II). For the other size groups of calanoid copepods, ZooImage had a tendency to overestimate the abundance (Fig. 4). From the confusion matrix (Table II), this was probably mainly due to misclassification within the copepod size groups. ZooImage had a tendency to underestimate *Oithona* spp., and overestimate *Oncaea* spp. (Fig. 4), probably mainly because of cases where the software was unable to differentiate between these two groups (Table II). Chaetognaths were overestimated (Fig. 4), probably mainly because some larvaceans were misclassified as chaetognaths (Table II).

Conclusions

Manual processing of zooplankton samples takes much time and labor. Thus, in our experience, a trained technician may analyse around one sample per day manually, whereas 6–12 samples may be processed by the automated methods using ZooImage, provided that the classifier is already made. This estimate includes the whole process, from the splitting of the samples with a Motoda splitter until the results on abundance and biomass of the groups are available in computerized form. By running the software in batch mode, the process may be speeded up even further. Other workers have reached a similar conclusion (e.g. Culverhouse, 2008).

While it is true that automated analysis of zooplankton samples occurs at the cost of loosing taxonomic information, there are several benefits that depending on the type of study may counter this limitation. Thus, the automated process gives information on parameters, such as size distribution and biomass, that are not easily obtained by the conventional methods, but are important parameters for understanding the structure and functioning of ecosystems. In addition, the automated methods enable the researcher to process many more samples than would be possible by the conventional methods, thus increasing the spatial and temporal resolution of the study. As noted earlier, even experts make mistakes when

manually sorting plankton samples, and during routine analysis by trained personnel the human error may even be as high as ~30% (Culverhouse *et al.*, 2003). However, while the human performance is likely to be variable and dependent on the analyser, that of the automated analysis is likely to be constant and known and thus easier to take into account when evaluating the results. Against this background, the new methods are a valuable addition to the conventional ones. It should be realized, however, that the conventional methods will remain, not only for comparative purposes with the automated ones, but also in studies that require identification to the species or developmental stage level, as for instance in studies of zooplankton community structure, development and life history patterns.

ACKNOWLEDGEMENTS

We are grateful to the crew on the RV "Bjarni Sæmundsson" and colleagues at the Marine Research Institute for their help in the sampling. We also want to thank Kevin Denis for valuable help and guidance in using the ZooImage software. Finally, we thank the anonymous referees for their helpful comments.

REFERENCES

Alcaraz, M., Saiz, E., Calbet, A. *et al.* (2003) Estimating zooplankton biomass through image analysis. *Mar. Biol.*, **143**, 307–315.

Bell, J. and Hopcroft, R. (2008) Assessment of ZooImage as a tool for the classification of zooplankton. *J. Plankton Res.*, **30**, 1351–1367.

Benfield, M. C., Grosjean, P., Culverhouse, P. F. *et al.* (2007) RAPID research on automated plankton identification. *Oceanography*, **20**, 172–187.

Breiman, L. (2001) Random forests. *Mach. Learn.*, **45**, 5–32.

Culverhouse, P. (2008) Automatic visual identification of plankton. *Report of SCOR WG130*, 31 pp.

Culverhouse, F. P., Williams, R., Reguera, B. *et al.* (2003) Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. *Mar. Ecol. Prog. Ser.*, **247**, 17–25.

Davis, C. S., Hu, Q., Gallager, S. M. *et al.* (2004) Real-time observation of taxa-specific plankton distributions: an optical sampling method. *Mar. Ecol. Prog. Ser.*, **284**, 77–96.

Estep, K. E., MacIntyre, F., Hjörleifsson, E. *et al.* (1986) MacImage: a user-friendly image-analysis system for the accurate mensuration of marine organisms. *Mar. Ecol. Prog. Ser.*, **33**, 243–253.

Fernandes, J. A., Irigoien, X., Boyra, G. *et al.* (2009) Optimizing the number of classes in automated zooplanton classification. *J. Plankton Res.*, **31**, 19–29.

Gashler, M., Giraud-Carrier, C. and Martinez, T. (2008) Decision tree ensemble: small heterogeneous is better than large homogeneous. Seventh International Conference on Machine Learning and Applications (ICMLA 08), 900–905, doi:10.1109/ICMLA.2008.154.

Grosjean, P. and Denis, K. (2007) Zoo/PhytoImage version 1.2-0. *User's Manual.* 57 pp.

Grosjean, P., Picheral, M., Warembourg, C. *et al.* (2004) Enumeration, measurement, and identification of net zooplankton samples using the ZOOSCAN digital imaging system. *ICES J. Mar. Sci.*, **61**, 518–525.

Hu, Q. and Davis, C. (2006) Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. *Mar. Ecol. Prog. Ser.*, **306**, 51–61.

Irigoien, X., Grosjean, P. and Urrutia, A. L. (2005) Image analysis to count and identify zooplankton. GLOBEC/SPACC Workshop, San Sebastian, November2005 21 pp.

Irigoien, X., Fernandes, J. A., Grosjean, P. *et al.* (2009) Spring zooplankton distribution in the Bay of Biscay from 1998 to 2006 in relation with Anchovy recruitment. *J. Plankton Res.*, **31**, 1–17.

Kouzani, A. Z., Nahavandi, S. and Khoshmanesh, K. (2007) Face classification by a random forest. TENCON 2007 - 2007 IEEE Region 10 Conference, pp. 1–4.

Motoda, S. (1959) Devices of simple plankton apparatus. *Mem. Fac. Fish. Hokkaido Univ.*, **7**, 73–94.

Ortner, P. B., Cummings, S. R., Aftring, R. P. *et al.* (1979) Silhouette photography of oceanic zooplankton. *Nature*, **277**, 50–51.

Plourde, S., Joly, P. and Irigoien, X. (2008) A preliminary assessment of the performance of an automated system for the analysis of zooplankton samples from the Gulf of St. Lawrence, Nortwest Atlantic. *AZMP, Bull.*, **7**, 42–47.

Rolke, M. and Lenz, J. (1984) Size structure analysis of zooplankton samples by means of an automated image analysis system. *J. Plankton Res.*, **6**, 637–645.

Shi, T., Seligson, D., Belldegrun, A. S. *et al.* (2005) Tumor classification by tissue microarray profiling: random forest clustering applied to renal cell carcinoma. *Modern Pathol.*, **18**, 547–557.

Simpson, R., Williams, R., Ellis, R. *et al.* (1992) Biological pattern recognition by neural network. *Mar. Ecol. Prog. Ser.*, **79**, 303–308.

Zarauz, L., Irigoien, X., Urtizberea, A. *et al.* (2007) Mapping plankton distribution in the Bay of Biscay during three consecutive spring surveys. *Mar. Ecol. Prog. Ser.*, **345**, 27–39.