JOURNAL OF PLANKTON RESEARCH | VOLUME 30 | NUMBER 12 | PAGES 1351–1367 | 2008

Assessment of ZooImage as a tool for the classification of zooplankton

JENEFER L. BELL AND RUSSELL R. HOPCROFT*
INSTITUTE OF MARINE SCIENCE, UNIVERSITY OF ALASKA, FAIRBANKS, AK 99775-7220, USA

*CORRESPONDING AUTHOR: hopcroft@ims.uaf.edu

Received July 24, 2008; accepted in principle September 10, 2008; accepted for publication September 14, 2008; published online September 15, 2008

Corresponding editor: Mark J. Gibbons

ZooImage, image analysis software, was evaluated to determine its ability to differentiate between zooplankton groups in preserved zooplankton samples collected in Prince William Sound, Alaska. A training set of 53 categories were established to train the software for automatic recognition. Using the Random forest algorithm, ZooImage identified particles in the training set with less than 13% error. Despite reasonable results with the training set, however, ZooImage was less effective when this training set was used to identify particles from field-collected zooplankton samples. When all particles were examined, ZooImage had an accuracy of 81.7% but this dropped to 63.3% when discard particles (e.g. marine snow and fibers) were removed from total particles. Copepods, the numerically dominant organisms in most samples, were examined separately and were correctly identified 67.8% of the time. Further investigation suggested size was effective in determining identifications; medium size copepods (e.g. Pseudocalanus sp., Acartia sp.) were accurately identified 73.3% of the time. ZooImage can provide a coarse level of taxonomic classification and we anticipate continued improvement to this software should further enhance automatic identification of preserved zooplankton samples.

INTRODUCTION

Ubiquitous throughout the world's oceans, zooplankton serves as the major prey for many higher trophic level organisms (e.g. Castro-Longoria *et al.*, 2001). Zooplankton communities vary both temporally and spatially, thus extensive effort must be undertaken to describe their broader-scale and seasonal patterns. Once baselines are established, zooplankton communities can become sensitive indicators of environmental conditions (e.g. Marine Zooplankton Colloquium 2, 2001; Benfield *et al.*, 2007). Thus, long-term observations on planktonic communities may be the key to understanding how marine environments, and the higher trophic levels dependent on them, change over time (Richardson, 2008).

Despite the existence of modern *in situ* technologies (see reviews by Sameota *et al.*, 2000; Wiebe and Benfield, 2003), plankton net systems remain the most common device for assessing zooplankton abundance. Collecting zooplankton with nets is a quick and inexpensive process, and most studies typically generate a large number of samples to be analyzed. Identifying marine zooplankton is, however, a significantly more time-consuming task that requires considerable taxonomic proficiency to generate accurate results (Ellis *et al.*, 1994; Zavala-Hamz *et al.*, 1996; Tang *et al.*, 1998). Ultimately, the constraint in understanding the temporal and spatial variations of zooplankton populations becomes not the simple collection of samples, but the processing effort required to yield basic information on abundance, biomass and species composition. These limitations have generated increasing interest in the development of methods to automatically identify zooplankton, both *in situ* and preserved (Culverhouse *et al.*, 2006).

Image analysis of plankton for identification began in earnest in the early 1980s with attempts to obtain measurements of particles and utilize discriminant analysis to identify them. Some early image analysis

doi:10.1093/plankt/fbn092, available online at www.plankt.oxfordjournals.org

systems did little more than accurately measure and enumerate particles (Rolke and Lenz, 1984; Estep *et al.*, 1986). Taking it one step further by using pattern recognition, others achieved 89% accuracy identifying organisms into eight taxonomic groups (Jeffries *et al.*, 1984). By the early 1990s, imaging techniques were improving but a real advance was the use of neural networks. Neural networks are a good choice for identification problems because they can be taught the correct answer, are not as readily affected by imperfect images, and are designed to run complex tasks at high speeds (Simpson *et al.*, 1992). A successfully trained neural network was capable of identifying five similar species of tintinnids (Culverhouse *et al.*, 1994). Despite some success, early systems were not effective in practical situations because they were quickly overwhelmed by poor imaging and slow computational ability. Additionally, the early systems did not handle large particle numbers and were restricted to lab prepared samples (Hu and Davis, 2005).

To some extent, early image analysis systems could only accommodate images that the user arranged to create the perfect image. The geometric regularity of phytoplankton made them a good candidate for image analysis (Gorsky *et al.*, 1989). Unlike phytoplankton, however, zooplankton present significant identification challenges because they come in a variety of sizes and morphologies, with significant distinction between groups (e.g. euphausiids and copepods), but often great similarity between related genera (e.g. *Neocalanus* spp. and *Calanus* spp.) (Alvarez-Borrego and Castro-Longoria, 2003). Not only could the same individual zooplankter be preserved and oriented in a seemingly infinite number of postures, but its size, as well as other morphological attributes, changes during its ontological development (Grosjean *et al.*, 2004). The morphological diversity within zooplankton is further increased by the numerous benthic organisms that release meroplanktonic larval stages into the pelagic environment. Finally, identification of zooplankton is complicated by the routine presence of large numbers of non-biological particles and marine snow (Culverhouse *et al.*, 2006).

Image analysis of zooplankton can only be useful if it can effectively handle real "field" samples. That is, it has to be able to evaluate zooplankton into several taxonomic groups and genera, regardless of posture or orientation, and be able to distinguish them from marine snow and non-biological particles (e.g. debris and fibers). With recent advances in technology, the ability to obtain high resolution images with large particle counts has greatly increased, and rekindled the desire for automatic zooplankton identification systems. Such automatic recognition systems rely on training sets and machine learning algorithms to teach a computer to predict the identification of zooplankton within a sample (Hu and Davis, 2006), and have only become viable due to the ongoing evolution of imaging technologies in concert with increased processing power of computers. For zooplankton, much of the research has concentrated on *in situ* plankton collectors that are bundled with automatic recognition software packages as researchers acknowledged the need to process the large number of images created by these devices (e.g. Benfield *et al.*, 1996; Davis *et al.*, 1996; Samson *et al.*, 2001; Luo *et al.*, 2004; Luo *et al.*, 2005; Hu and Davis, 2006). Unfortunately, these software systems are designed for specific devices and are not easily compatible with other imaging systems (Benfield *et al.*, 2007). Additionally, *in situ* devices are expensive and present logistical constraints on their operation, so net systems remain the dominant technology utilized to collect plankton. A hybrid approach is to subject traditionally collected zooplankton to image analysis after collection.

ZooImage is a computer-assisted plankton image analysis software package in development for predicting taxonomic identification of preserved zooplankton samples (Grosjean and Denis, 2007, ZooImage User's Manual). ZooImage (currently a beta-release) is open source software bundled with Java-based ImageJ and R, statistical software. Through the development of plug-ins for ImageJ, ZooImage can be modified to meet the requirements of the user and accommodate many different imaging systems. At present, ZooImage is capable of accuracies of 70–80% when dealing with 10–20 taxonomic categories of zooplankton in some ecosystems (Benfield *et al.*, 2007).

The purpose of this project is to assess the abilities of ZooImage in the ecosystem of Prince William Sound, AK, where taxonomic diversity is relatively low, and both the biomass and composition of the zooplankton community has been linked to the success of local fisheries (Boldt and Haldorson, 2003). At present, although zooplankton is collected routinely, logistical and financial constraints limit the analysis of these samples to simple estimation of the displacement volume of the Prince William Sound zooplankton community. Determining the composition of zooplankton samples may provide better information about the prey field for juvenile pink salmon (*Oncorhynchus gorbuscha*), such as large copepods, euphausiids, and pteropods (Cooney *et al.*, 2001; Boldt and Haldorson, 2002; Armstrong *et al.*, 2005). Survival of juvenile pink salmon in Prince William Sound may be dependent in part on the release of hatchery-reared fish coincident with peak abundances of their main prey items. Using ZooImage to analyze zooplankton samples would provide real-time information that could, therefore, assist in maximizing

survival of juvenile salmon released from hatcheries. In this study, we compared taxonomically identified zooplankton samples with the results of the ZooImage system to examine whether ZooImage can be used to identify zooplankton with enough accuracy and taxonomic detail to be useful in determining taxonomic composition in zooplankton samples obtained from Prince William Sound.

MATERIALS AND METHODS

Hardware

Samples were scanned on an Epson Perfection 4990 Photo color flatbed scanner to which a $20 \times 25 \times 2.8$ cm acrylic box-frame was attached with silicone adhesive. The scanner was mounted to a platform that allowed it to be tipped so the sample could be emptied through a pour spout (made from a canted-neck tissue culture flask and hose) fitted into one end of the acrylic frame. Initial work determined that crisper images were determined when zooplankton was scanned directly on the scanner's glass, as opposed to elevated several millimeters as occurs when animals are in plastic trays/dishes. All image acquisition and processing was conducted using a 3.2 GHz computer with a Pentium IV processor and 4 GB of RAM.

ZooImage

ZooImage is designed as open source software so can be adapted to meet the requirements of the user, and many of the basic features can be reconfigured to maximize the performance. This project used the original version without modification, as we would expect this to be the typical default for most users. There are three sections to the ZooImage software system: (i) acquiring the image, (ii) automatic particle recognition and (iii) analysis of series for ecological and biological information (http://www.sciviews.org/Zoo/PhytoImage).

The first section of ZooImage is designed to import and process images. Using VueScan 8.3.23 Professional scanning software, images were scanned at a resolution of 2400 dpi and 16 bit grayscale (Appendix 1); the entire image was scanned in <20 min. ZooImage processed the images, created vignettes (snapshots) of all particles, extracted a variety of features from each vignette and recorded relevant information into a metadata file, a process taking 15–35 minutes, depending on the number of particles in the sample. Initial efforts to scan samples inside plastic trays placed onto the scanner made it difficult to prepare large samples for scanning. Ultimately, placing the zooplankton sample directly on the glass also decreased sample preparation time.

The second section of ZooImage teaches the software particle recognition. The vignettes and metadata were sorted into a training set; a folder tree coinciding with the category of taxonomic detail defined by the user. The training set was used to create a "classifier" utilizing six different machine learning algorithms bundled with ZooImage. Linear discriminant analysis (LDA) uses linear combinations of measurements to discriminate between groups. Recursive partitioning tree (RPT) creates decisional trees with each node discriminating between randomly selected variables. K nearest neighbor (KNN) and learning vector quantization (LVQ) determine groups by minimizing distances between training set particles and codebooks (a type of training set summary), respectively. Neural networks (NN) utilize a neural web with intermediary layers between the input layer (measurements) and the output layer (groups). Random forest (RF) creates several decisional trees with each decision node discriminating between two variables (for complete descriptions of the algorithms see: LDA—Hastie *et al.*, 1994; RPT and Random Forest—Breiman, 2001; KNN—Peters *et al.*, 2002; LVQ—Tang *et al.*, 1998; NN—Simpson *et al.*, 1992).

Each classifier was evaluated using a 10-fold cross validation confusion matrix to determine the error rate between manual and automatic recognition. Cross validation is a method that randomly divides the training set into 10 equal fractions. The learning phase uses nine fractions and predicts the test set (or tenth fraction). The process is repeated 10 times and results combined into a confusion matrix where the diagonal (from top-left to bottom-right) represents correct predictions made by the computer, and values off the diagonal represent errors in predictions (Grosjean and Denis, 2007, ZooImage User's Manual).

The third step in ZooImage creates "series". Series are defined by the user as specific combinations of zooplankton samples that assess ecological and biological patterns. ZooImage uses the processed images and the classifier to automatically identify all particles in the samples included in the series. Simultaneously, ZooImage calculates species abundances, biomasses and the size spectra of each sample and then records this information in a file exportable to Microsoft Excel, Matlab, etc. This process was completed in less than 5 min, depending on the number of samples contained in the series.

Creation of the training set

ZooImage relies on a training set to teach the computer to identify zooplankton. The training set was established

by selecting organisms from preserved zooplankton samples collected from Prince William Sound and the Gulf of Alaska. Undamaged organisms for the training set were chosen to represent typical individuals of the most common taxa in Prince William Sound with identification verified by an experienced taxonomist. For species that were generally abundant in the study area, 30 specimens of a taxon were selected, while we only succeeded in finding as few as 10 individuals for some of the rarer categories. The individuals of each taxon were scanned multiple times, with mixing of the sample's fluid and rearrangement between scans to maximize the number of orientations captured by the scanner. ZooImage was used to process the images and extract vignettes.

The training set generally followed the level of detail used by taxonomists conducting zooplankton studies in Prince William Sound; e.g. copepod taxa were evaluated to genus and stage, while other taxa were combined at coarser taxonomic levels (e.g. barnacle nauplii). The training set was developed using 100 vignettes per taxon that were selected to represent the standard condition and orientation of each organism. The training set also included categories for "discards", such as marine snow, non-biological particles (e.g. debris and fibers) and biological particles found in high abundances in zooplankton samples but not included in the taxonomic work (e.g. phytoplankton and calanoid nauplii). Discard categories were developed using a combination of vignettes from scanning the organisms (fibers were in all scans) and adding marine snow vignettes from several different sample scans. It is essential to include these discards in the training set so they can be classified correctly and removed before beginning analysis.

Finally, the training set started at the highest level of taxonomic detail possible and was reworked several times to maximize predictive ability while attempting to maintain a high level of taxonomic distinction. In some instances, it was beneficial to combine copepod stages or remove rarer species that were confounding the identification of more common species.

Validation samples

Taxonomically identified zooplankton samples were used for the comparison against ZooImage predictions. Plankton samples were collected in 1997 by Prince William Sound Aquaculture Corporation (Cordova, AK, USA) at a remote hatchery (Armin F. Koernig) in Prince William Sound using a 0.5 m diameter, 243 μm mesh ring-net towed vertically from 20 m depth to the surface. Plankton samples consisted of 1–3 pooled vertical tows. Taxonomic work on the plankton samples was completed in 1997 and samples were placed in storage until 2006. For ZooImage validation, 53 of the original 1997 plankton samples were utilized in the comparison.

Sample preparation

Samples were strained onto a 150 μm mesh screen and rinsed into a Folsom splitter using artificial seawater. Each sample was split to obtain the same fraction used in the earlier taxonomic analysis (typically 1/64–1/8). The fractionated sample was diluted to 250 mL with artificial seawater and placed into the frame for scanning. To improve accuracy in identification, particles were superficially dispersed to spread the sample and reduce the number of particles touching each other and the edges of the images. Splitting of a sample and preparing it for scanning were accomplished in under 10 min.

ZooImage limits the image size that can be imported (approx. 10 cm × 10 cm). This required the area containing the fractionated sample (approx. 19 cm × 23 cm) to be split into six smaller sub-images. BATCH mode in VueScan 8.3.23 was utilized to create the sub-images. All sub-images from a fractionated sample were processed individually and re-combined for automatic classification. ZooImage ignores particles touching the edges of the sub-images during vignette extraction, and consequently allows a percentage correction for this bias. We determined the "cell-part" correction for fractionated samples by counting all particles touching the edges of randomly selected sub-images of 26 different fractionated samples. The total number of edge particles was divided by the total number of particles in each sub-image and then the average for the 26 samples was obtained.

Determining ZooImage accuracy

The accuracy of ZooImage automatic recognition was evaluated by comparing the computer prediction against the manual identification of vignettes from 20 of the 53 validation samples. Discard categories (i.e. debris, fiber, marine snow, calanoid nauplii and phytoplankton) made up the majority of the particles and were labeled as DTP (discard true positive) if the particle was correctly identified into a discard category. If the computer identification was incorrect there were two options: DFP (discard false positive) if the particle was identified as a discard but was actually something else or DFN (discard false negative) if the particle was identified as something outside the discard categories. Generally, copepods were the next most abundant groups, so determining the

accuracy in prediction of copepods was similar: CTP (copepod true positive) if the particle was correctly identified as a copepod, CFP (copepod false positive) when the computer identified a particle as a copepod and it was not, and CFN (copepod false negative) when a particle was actually a copepod but was identified as something other than a copepod. If a particle was labeled CTP, then it was further evaluated to determine if the computer correctly identified it to group (e.g. copepod) or had the right genus but the wrong stage (e.g. *Pseudocalanus* copepodite I versus *Pseudocalanus* copepodite II). For additional analysis, copepods were separated into size classes using established measurements from taxonomists working in Prince William Sound: small (<0.9 mm), medium (0.91–1.50 mm) and large (>1.51 mm). Particles in all other categories were labeled C (correct) or I (incorrect) or NC (not classified) for particles that were misidentified because they were organisms not represented in the training set. The error associated with automatic recognition fell into two categories: user error and machine error. User error was defined as misidentifications caused by sample set-up (e.g. bad images or touching particles). Machine error was defined as misidentifications due to inconsistencies in particle identification.

RESULTS

Training set

Based on the ongoing taxonomic work conducted in Prince William Sound, the training set was established with 63 categories that accurately reflected zooplankton community composition (Table I). The classifiers created from the initial training set using ZooImage's six different algorithms showed large differences in the k-fold cross validation error, ranging from 63 to 15% (Table II). Random forest was selected for the continued work because it produced the lowest k-fold cross validation error. The training set was established by combining and removing individual taxa to minimize the

Table I: Categories used to create training sets

Original Set	Final Set	Original Set	Final Set
Amphipoda	Amphipoda	*N. plum/flem* II	*N. plum/flem* II
Copepoda	Copepoda	*N. plum/flem* III	*N.plum/flem* III
Acartia II	*Acartia* II	*N. plum/flem* IV	*N. plum/flem* IV
Acartia III	*Acartia* III,IV,V,F,M	*N. plumchrus* V	*N. plum/flem* V
Acartia IV		*N. flemengeri* V	
Acartia V		*Oithona*	*Oithona*
Acartia Female		*Pseudocalanus* I	*Pseudocalanus* I
Acartia Male		*Pseudocalanus* II	*Pseudocalanus* II
Calanoid nauplii	Calanoid nauplii	*Pseudocalanus* III	*Pseudocalanus* III
Calanus III	*Calanus* III	*Pseudocalanus* IV	*Pseudocalanus*
Calanus IV	*Calanus* IV	*Pseudocalanus* V F	IV, V, F, M
Calanus V	*Calanus* V	*Pseudocalanus* V M	
Calanus VI	*Calanus* VI	Chaetognatha	Chaetognatha
Centropages IV		Cirripedia cyprids	Cirripedia cyprids
Centropages V		Cirripedia nauplii	Cirripedia nauplii
Centropages Female		Cnidaria	Cnidaria
Centropages Male		Cyphonautes	Cyphonautes
Eucalanus nauplii	*Eucalanus* nauplii	Discards	Discards
Eucalanus I	*Eucalanus* I	debris	debris
Eucalanus II	*Eucalanus* II	fiber	fiber
Eucalanus III	*Eucalanus* III	marine snow	marine snow
Eucalanus IV	*Eucalanus* IV	Euphausiacea	Euphausiacea
Eucalanus V	*Eucalanus* V	calytopids	calytopids
Metridia I	*Metridia* I	eggs	eggs
Metridia II	*Metridia* II	furcillids	furcillids
Metridia III	*Metridia* III	metanauplii	metanauplii
Metridia IV	*Metridia* IV	nauplii	nauplii
Metridia V Female	*Metridia* F, M	Larvaceans	Larvaceans
Metridia V Male		*Fritillaria*	*Fritillaria*
Neocalanus cristatus III	*N. cristatus* III	*Oikopleura*	*Oikopleura*
N. cristatus IV	*N. cristatus* IV	Phytoplankton	Phytoplankton
N. cristatus V	*N. cristatus V*	Polychaete larva	Polychaete larva
N. plumchrus/flemengeri I	*N. plum/flem* I	Shrimp	Shrimp
		Thaliacea	Thaliacea

The original training set contained 63 categories; the final training set contained 53 categories.

Table II: The six machine algorithms bundled with ZooImage and their k-fold cross validation errors based on a 63 category training set

Algorithm	k-fold cross validation error (%)
Linear discriminant analysis	41.3
Recursive partitioning tree	63.4
K nearest neighbor	51.6
Learning vector quantization	59.0
Neural network	35.5
Random forest[a]	15.2

[a]This algorithm was utilized in evaluating ZooImage.

k-fold cross validation error. The training set was reduced to 53 categories (Fig. 1) with a final k-fold cross validation = 12.54% (0–36% for each taxon).

Validation samples

ZooImage results were compared to taxonomically identified samples, in order to determine the ability of ZooImage to detect patterns in field-collected samples. Prior to the analysis, ZooImage and taxonomic abundances were log-transformed to satisfy tests for heteroscedasticity and normalacy (SigmaPlot v11). Additionally, a one to one relationship was assumed so the regression was forced

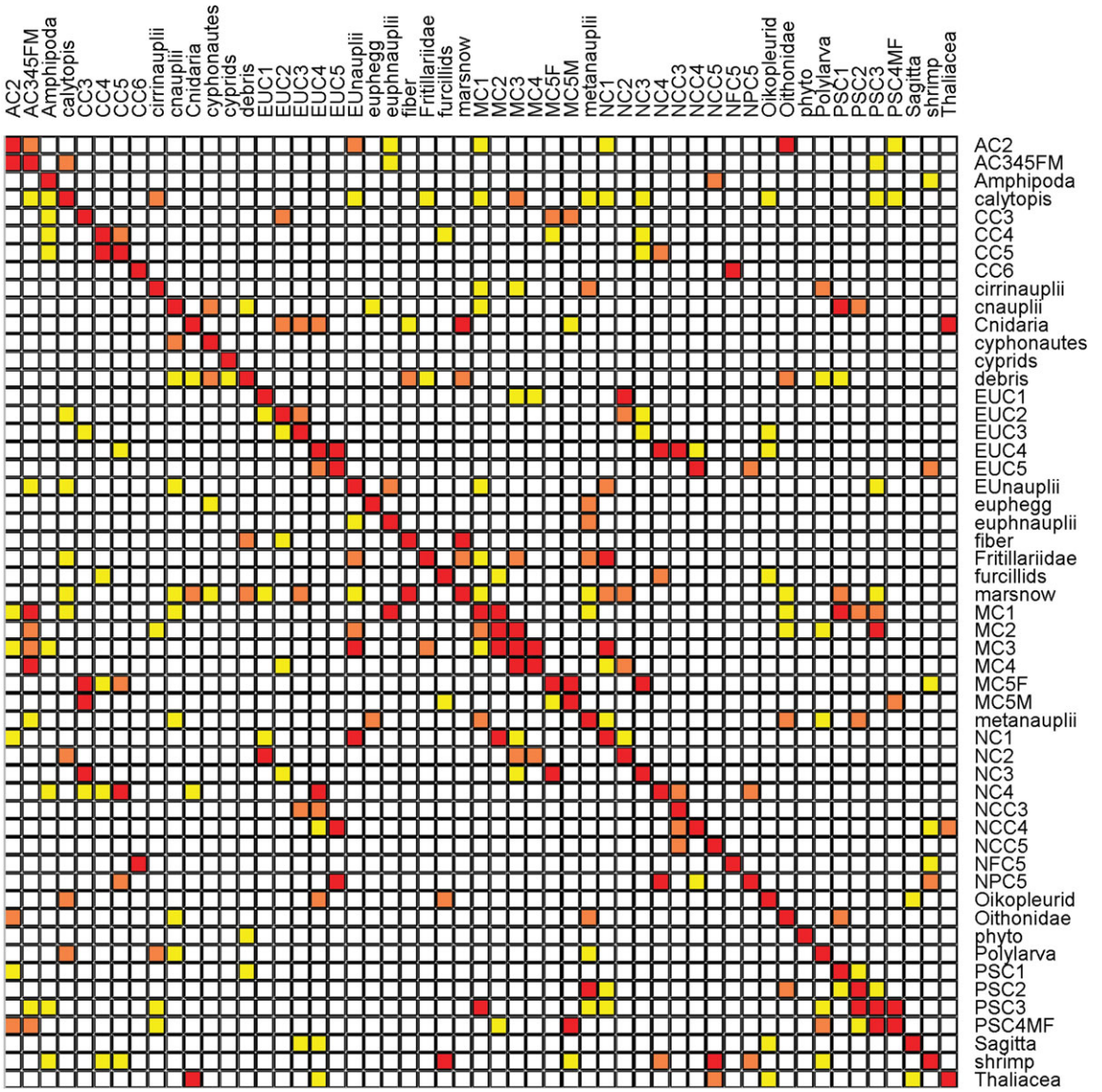

Fig. 1. The Confusion matrix for the 53 categories of the training set using the random forest algorithm. Rows are taxonomic classification and columns are automatic recognition of the same classifications. The diagonal represents correct identifications and points outside the diagonal represent errors in classification; the darker the color, the higher the value in the cell.

through zero. When total number of biological particles was assessed, ZooImage did a reasonable job of detecting those particles. Only one sample (April 4), which was dominated by small barnacle nauplii, was misidentified; removing this point from the analysis yielded a better relationship between ZooImage and taxonomic abundances (Fig. 2). Copepods were the dominant biological particles in most samples and ZooImage was more effective at capturing the trends in copepod abundance (Fig. 3). Comparing size classes of copepods between taxonomic and ZooImage results demonstrated that ZooImage was not as reliable in detecting small copepods (Fig. 4). In contrast, ZooImage was very successful at classifying medium-sized copepods (Fig. 5), which was the most abundant size class in most samples. Euphausiid eggs were the only other group present in all samples in great

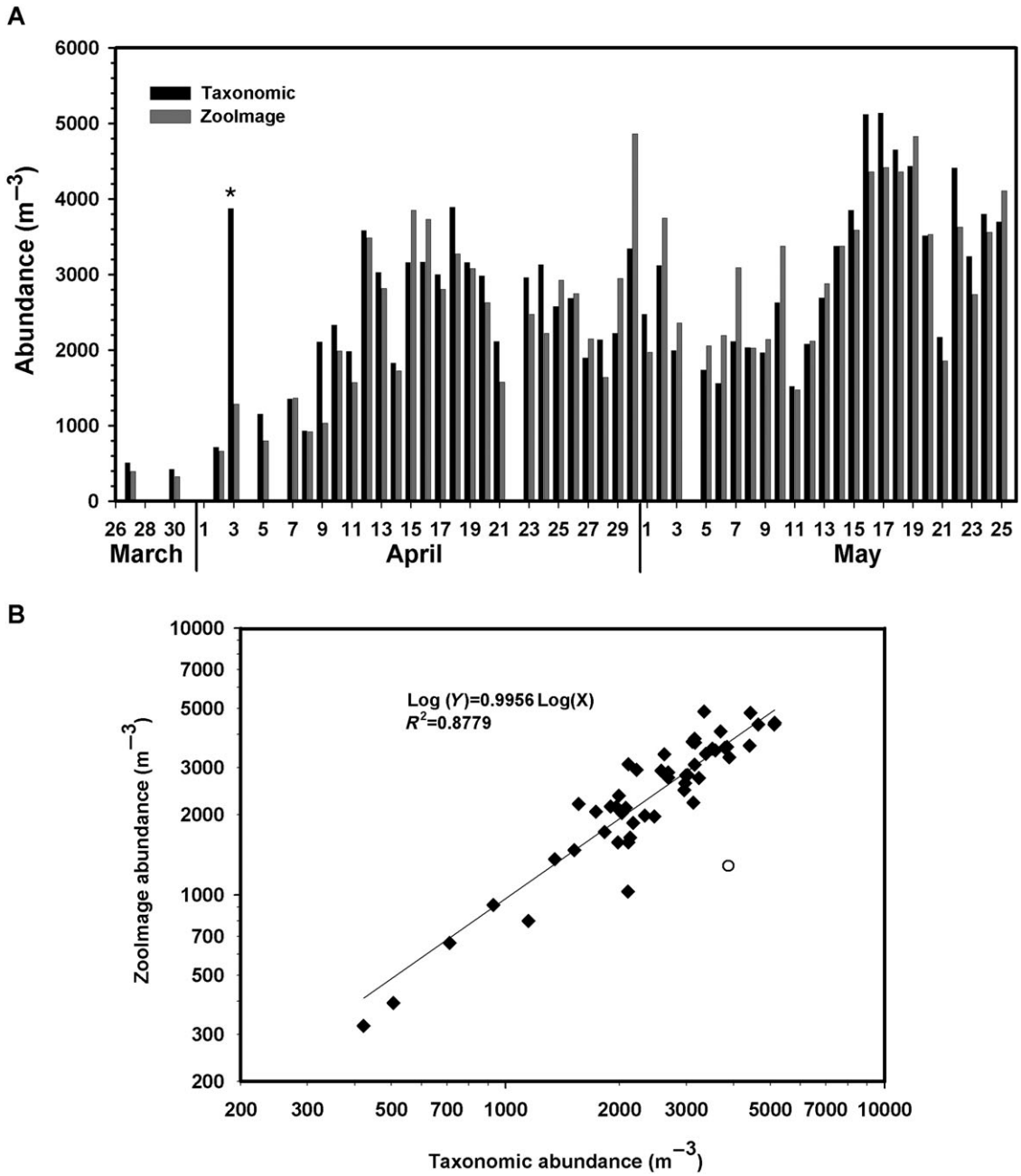

Fig. 2. Comparison of ZooImage and taxonomic abundances of biological particles. (**A**) Comparison of trends between ZooImage and taxonomic abundances; *indicates a taxonomic sample dominated by small barnacle nauplii that were not effectively identified by ZooImage. (**B**) Relationship between taxonomic and ZooImage abundances of biological particles. ○ indicates a taxonomic sample dominated by small barnacle nauplii; this point was removed from analysis.

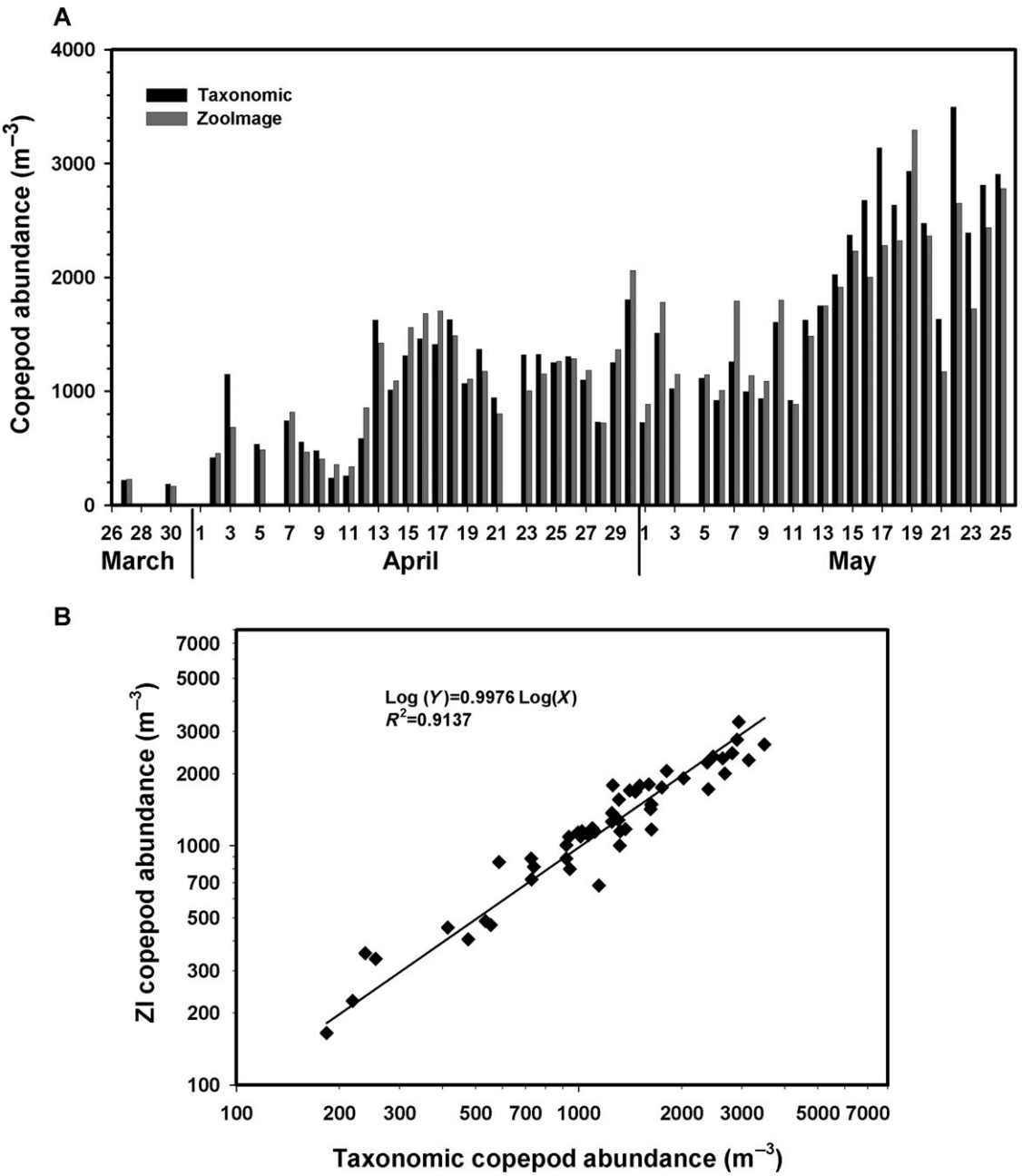

Fig. 3. Comparison of ZooImage and taxonomic copepod abundances. (**A**) Comparison of trends between ZooImage and taxonomic copepod abundances. (**B**) Relationship between taxonomic and ZooImage copepod abundances.

enough numbers to make reasonable comparisons. Unfortunately, ZooImage was not effective at identifying euphausiid eggs. The poor fit of this relationship was driven by one sample, in which ZooImage was unable to correctly identify euphausiid eggs; removing this sample point substantially improved the relationship (Fig. 6). All other groups did not occur in a sufficiently large number of samples or at a high enough abundance to make effective comparisons; this included large copepods.

ZooImage accuracy

ZooImage accuracy was assessed by comparing automatic recognition and actual identification of all particles in 20 samples. Discards were abundant and accounted for more than 75% of particles in all samples. Copepods had the second highest abundance in most samples but were surpassed in several samples by euphausiid eggs. The top five categories were

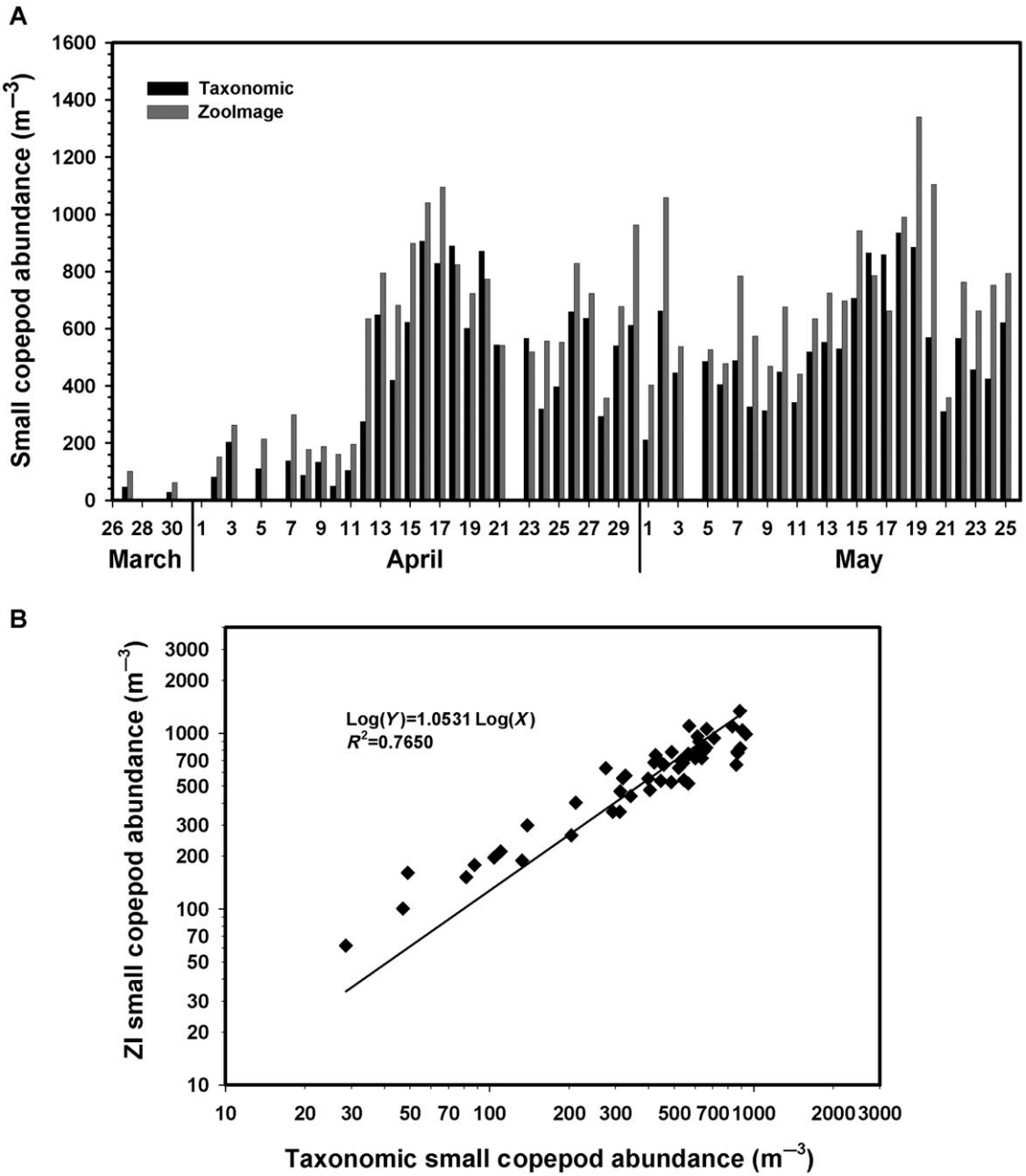

Fig. 4. Comparison of ZooImage and taxonomic small copepod abundances. (**A**) Comparison of trends between ZooImage and taxonomic small copepod abundances. (**B**) Relationship between taxonomic and ZooImage small copepod abundances.

relatively consistent between samples; copepods, euphausiid eggs, euphausiid nauplii, bryozoa larvae and barnacle nauplii were common in the early April samples, while meta-nauplii replaced barnacle nauplii in May samples (Fig. 7).

ZooImage correctly identified particles 81.7% of the time, with 18.3% error arising from 2.3% user error and 16.0% machine error (Fig. 8). Discard categories accounted for less than 5% of the machine error (DFN=3.9%, DFP=1.0%). When discards were removed, ZooImage correctly identified the remaining particles 63.3% of the time with 31.3% machine error and 5.4% user error (Fig. 8). Further assessment of the samples showed copepods were correctly identified 67.8% of the time with user error of 4.7%, CFN of 12.5% and CFP of 15.4% (Fig. 9).

While copepods were the numerically dominant taxon in most samples, the level of taxonomic detail

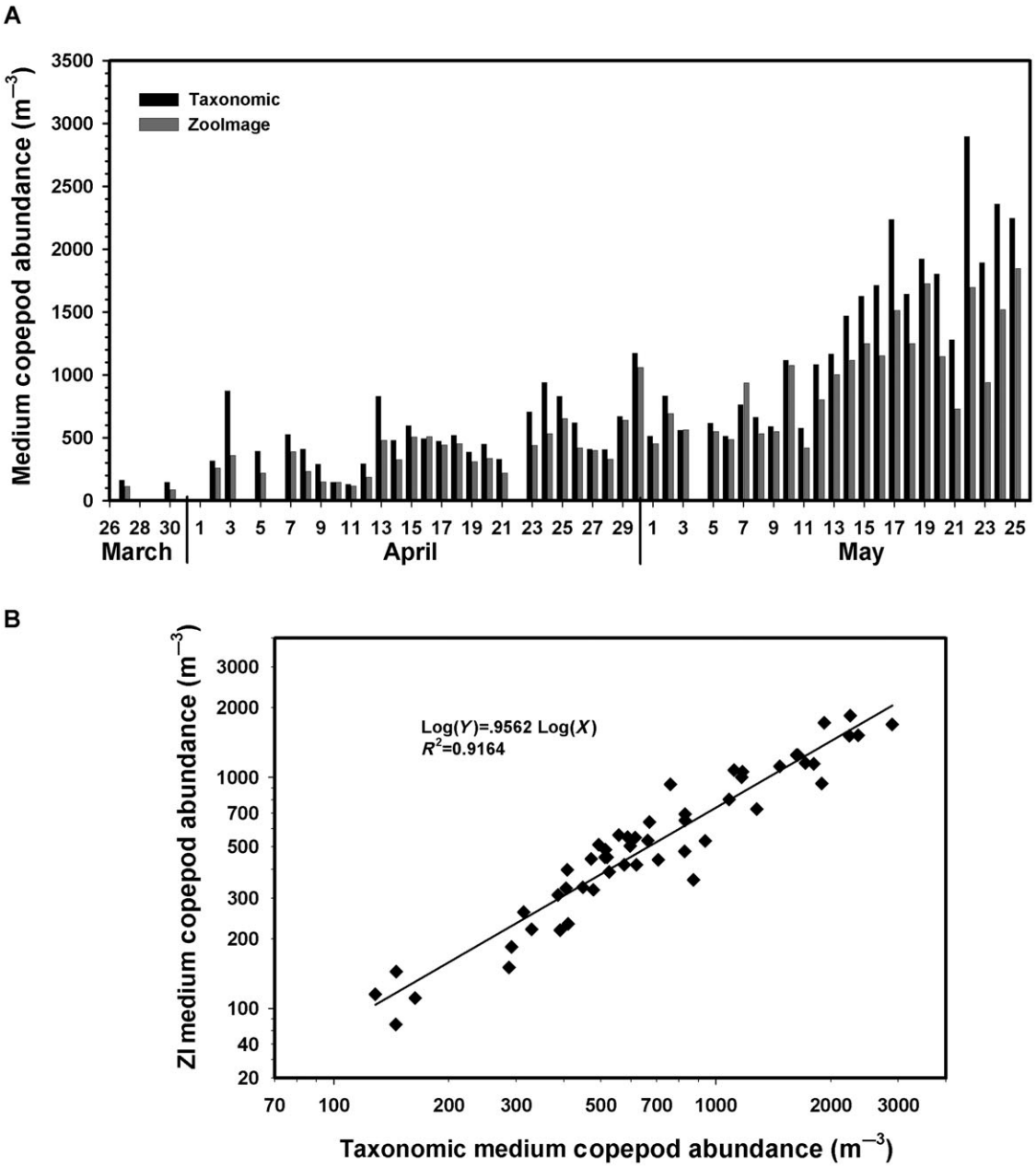

Fig. 5. Comparison of ZooImage and taxonomic medium copepod abundance. (**A**) Comparison of trends between ZooImage and taxonomic copepod abundances. (**B**) Relationship between taxonomic and ZooImage medium copepod abundances.

expected proved challenging for automatic recognition making the assessment of accuracy to genus level difficult. For example, ZooImage, in general, correctly classified particles as *Pseudocalanus* sp. but incorrectly identified *Acartia* sp. as *Pseudocalanus* sp. as well. Because of this, accuracy of automatic recognition was evaluated based on size classes (small, medium and large) of copepods. ZooImage only correctly identified large copepods with 59.4% correct identification and 40.6% errors (CFN = 3.3%, CFP = 29.6%, user = 7.7%) (Fig. 9). ZooImage achieved slightly better results when identifying small copepods, with correct identification of 62.5% and errors of 37.5% (CFN = 10.2%, CFP = 23.7%, user = 3.6%) (Fig. 9). ZooImage achieved the best results when identifying medium copepods; correct identification = 73.3% with errors = 26.7% (CFN = 14.6%, CFP = 7.3%, user = 4.8%) (Fig. 9).

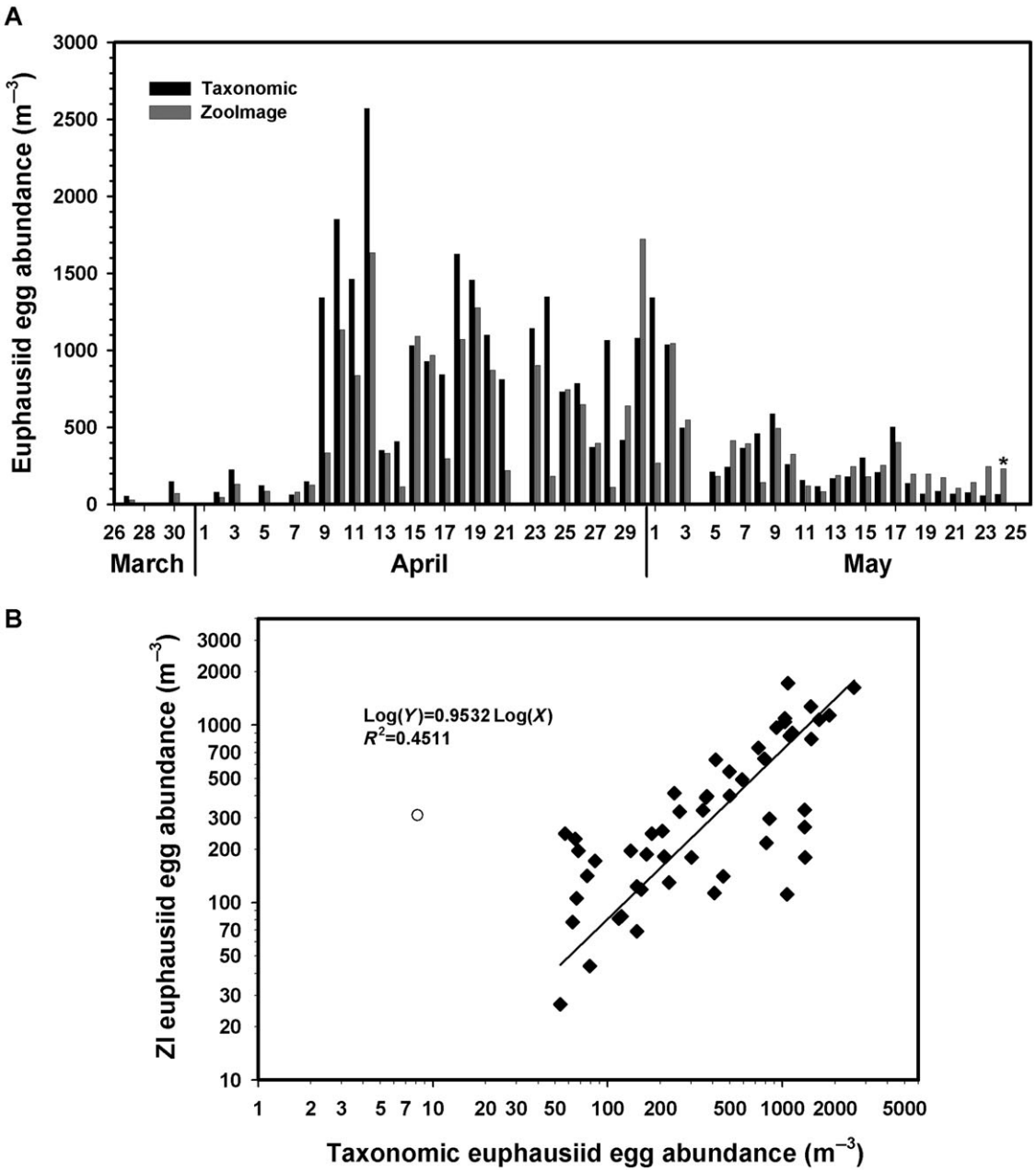

Fig. 6. Comparison of ZooImage and taxonomic abundances of euphausiid eggs. (**A**) Comparison of trends between ZooImage and taxonomic abundances of euphausiid eggs. * indicates a sample where ZooImage incorrectly identified euphausiid eggs. (**B**) Relationship between taxonomic and ZooImage euphausiid egg abundance. ○ indicates a sample where ZooImage incorrectly identified euphausiid eggs and was removed from the analysis.

DISCUSSION

At present, it appears unlikely that the image analysis will completely replace manual processing of zooplankton samples in the near future. Consequently, sample processing will remain the bottleneck (Ellis *et al.*, 1994) in large-scale projects. In many cases, however, a coarser level of taxonomic detail may be sufficient and image analysis could provide an appropriate level of discrimination. Image analysis systems can only be effective if they possess certain characteristics. The identification of zooplankton must be achieved with enough accuracy to draw conclusions about community composition. Additionally, image analysis systems should require limited technical expertise and accelerate

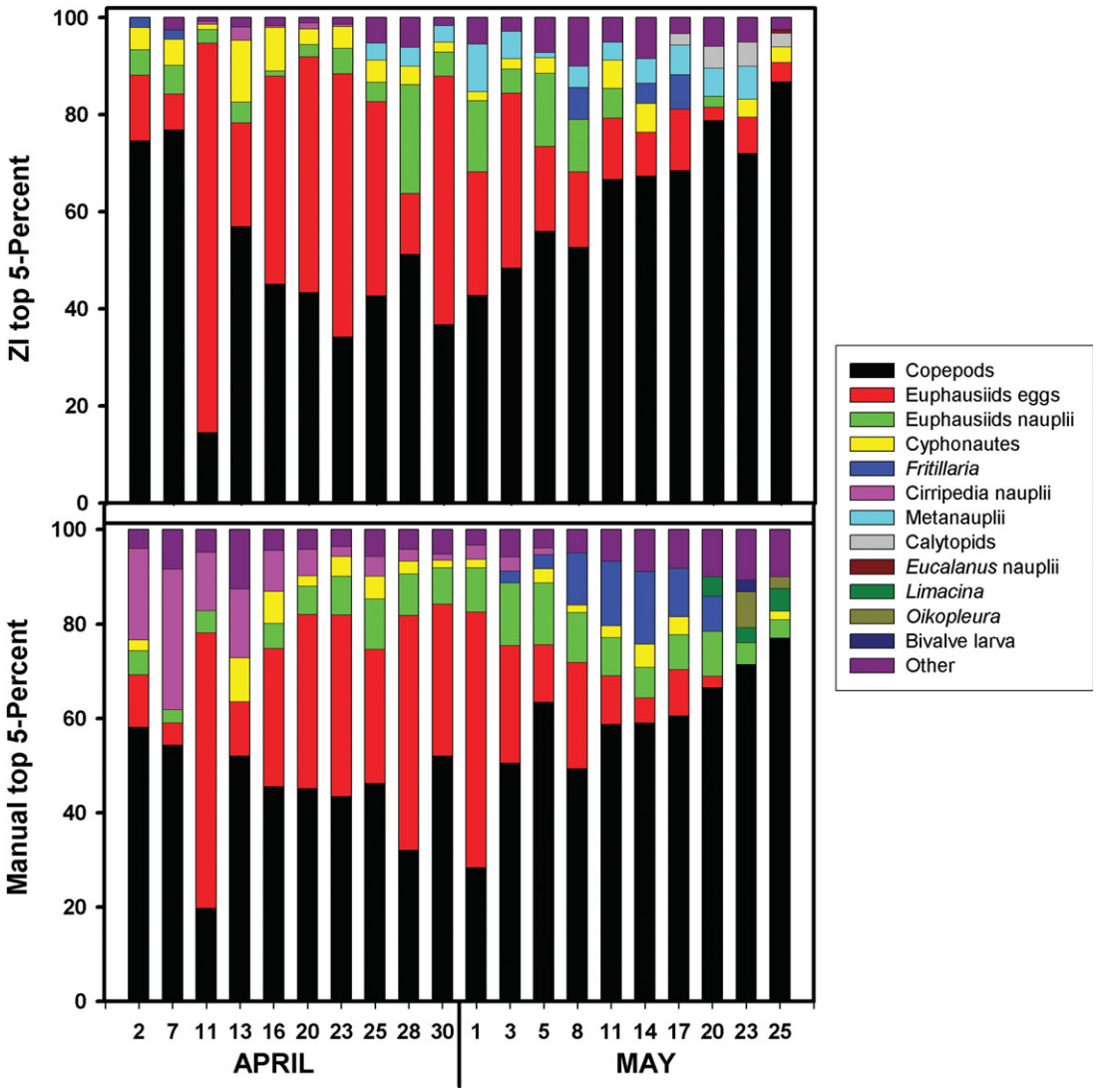

Fig. 7. Comparison of top five most abundant zooplankton categories identified by ZooImage (top) and manually (bottom). Based on the number of correct identifications; "other" combines all groups outside the top five.

zooplankton sample processing (Gorsky *et al.*, 1989; Culverhouse *et al.*, 2006). ZooImage demonstrates these attributes for examining preserved zooplankton samples.

The ability of ZooImage to correctly identify particles is directly related to how well the training set represents and encompasses the zooplankton present in the samples to be analyzed (Culverhouse *et al.*, 1994; Embleton *et al.*, 2003). Potentially, refining both the training set and user setting of the software might yield further improvement in accuracy. The proximal goal of our effort, however, was simply to ascertain the typical level of taxonomic classification possible by routine users of ZooImage, based on a training set developed with a moderate amount of effort, that was then applied to field-collected zooplankton samples in Prince William Sound. While studies have focused on discriminating phytoplankton species (Ishii *et al.*, 1987, Estep and MacIntyre, 1989; Gorsky *et al.*, 1989; Embleton *et al.*, 2003; Hense *et al.*, 2008), and microzooplankton (Culverhouse *et al.*, 1994) using image analysis, little work has been conducted on preserved zooplankton samples (Samson *et al.*, 2001). Although, ZooImage (version 1.2-1) is still in development, as are the techniques associated with its application, it does show promise as a rapid, yet effective, image analysis/recognition system for examining zooplankton composition.

In this study, "Random Forest" was used to develop the training set in a highly iterative process that required consolidation and/or removal of the initial 63 categories to yield the highest overall accuracy in

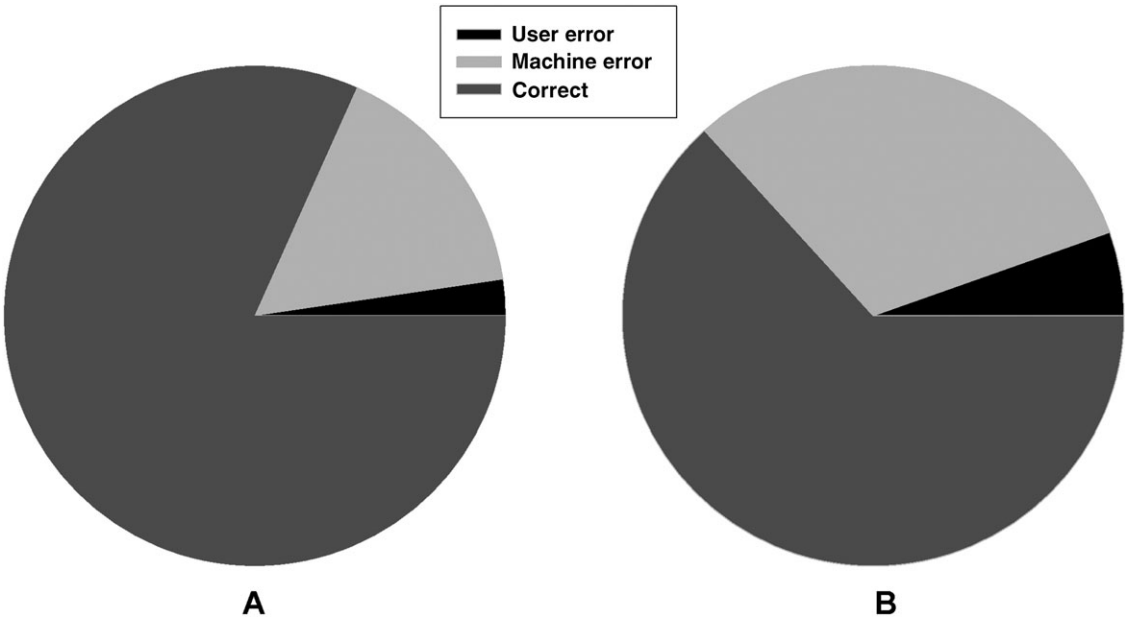

Fig. 8. Accuracy of ZooImage identification in 20 fractionated samples. (**A**) All particles. (**B**) Biological particles.

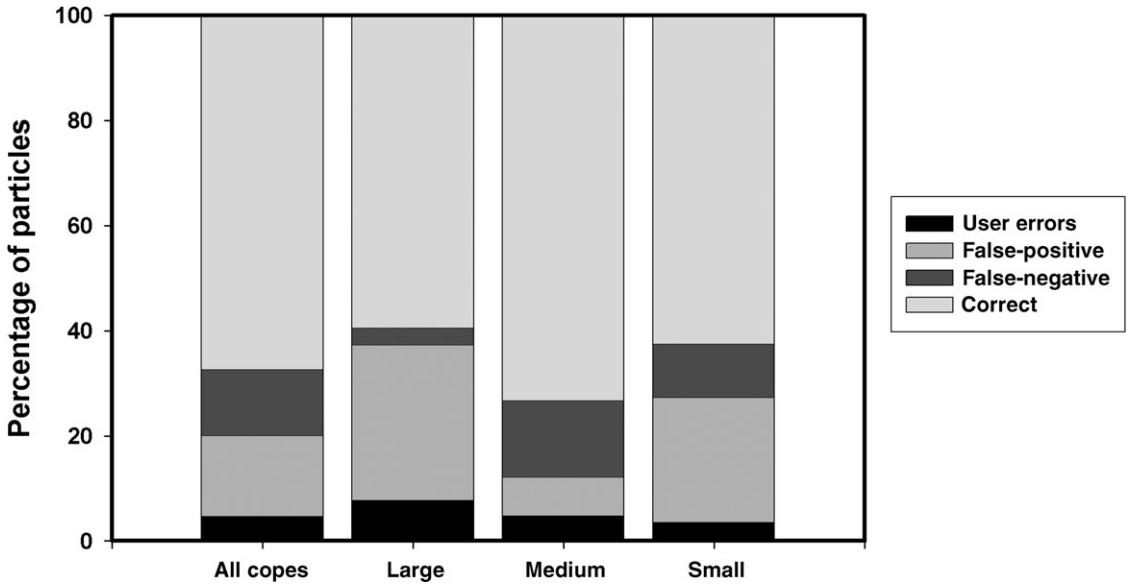

Fig. 9. Accuracy of ZooImage's identification of copepods with respect to size, and the associated user and machine errors.

identification with the highest level of taxonomic detail, both in the training set and when applied to real samples. The final training set of 53 categories (Table I) was designed to obtain the smallest k-fold cross validation error but maintain a high level of taxonomic distinction so field collected samples could be effectively analyzed. The number of categories used in this study exceeds that used in other studies (Grosjean *et al.*, 2004; Benfield *et al.*, 2007) and despite encouraging results from the training set (i.e. >85% correct identifications), results were less definitive when ZooImage was used for automatic recognition of zooplankton communities within field collections.

There may be several explanations for the disparity between the training set results and those obtained from the preserved samples. Differences can arise because the training set was built with only ideal images of particles of interest, whereas in field samples the computer has to classify all particles, including numerous discard categories and those particle images that would be

challenging even for an expert to classify (Hu and Davis, 2006). A taxonomist analyzing samples also has the ability to re-orient zooplankters to facilitate accurate identifications, while automated systems are constrained by the orientation of the scanned particles (Culverhouse *et al*., 2003). Finally, the technique for creating the training set used in this study may have contributed to some of the differences in accuracy between the training set and the samples. Using a limited number of organisms and incorporating different views of those organisms may not have provided sufficient baseline information about the groups to accurately identify them in field samples. A larger number of organisms, each represented by a single image, are suggested for future applications.

The accuracy of ZooImage was assessed by comparing the computer identification of each particle to the visual identification of each particle image. Evaluating the accuracy of ZooImage identification is challenging because it has been suggested that even experts can, and do, make mistakes when identifying plankton (Culverhouse *et al*., 2003). However, the diversity of zooplankton in Prince William Sound is limited and this, coupled with the established categories (e.g. combining *Pseudocalanus* sp. IV, V, M and F), may reduce errors associated with mistaken staging which is likely to be more prevalent than mistaken species identification.

When all particles were incorporated, ZooImage had greater than 80% accuracy when identifying all particles in a sample. Discards made up only 5% of the remaining errors despite being the dominant group in all samples. This suggests that ZooImage does a reasonable job detecting ill-defined particles that are difficult to standardize. These results are consistent with the analysis of *in situ* devices where accuracy dropped from 90 to 75% when unidentifiable particles were part of the data set (Luo *et al*., 2003). Copepods made up the majority of biological particles in most samples, and typically contain high diversity. While, 67.8% of the biological particles were correctly identified as copepods, ZooImage was less effective at separating copepods into genera. There has been limited success using diffraction patterns to obtain differentiation between two genera of copepods, *Calanus* sp. and *Acartia* sp. (Zavala-Hamz *et al*., 1996), between five different copepod genera (Castro-Longoria *et al*., 2001) and four *Acartia* species (Alvarez-Borrego and Castro-Longoria, 2003). While the training set analysis suggested an ability to identify several different genera, for field samples the system was only proficient at classifying a few of these categories, suggesting that size rather than taxonomy might be a better characteristic for achieving improved accuracy.

It appears that ZooImage's machine learning algorithms relied heavily on length and equivalent circular diameter (ECD) to discriminate between organisms within our training set. There were few incidents where organisms that are normally of different sizes were confused; e.g. *Calanus* sp. and *Pseudocalanus* sp., despite similar body shapes were rarely confused. The image analysis extracts numerous values from the images, including ECD and attributes such as optical density; however, it is the classification algorithm that decides which of those indices contribute to the classification. The algorithms used by ZooImage do not rely on ECD by default, but indicate that ECD was one of the most informative parameters to distinguish the categories in our training set and its consequent application to field samples.

Nonetheless, the use of size also introduced problems when identifying large copepods. Essentially, large copepods (e.g. *Neocalanus* sp., *Calanus* sp. and *Eucalanus* sp.) were the biggest items in the training set, so usually large non-biological particles and marine snow (not represented in the training set) were generally misclassified as large copepods. This problem could be minimized by removing large "discard" particles prior to scanning, eliminating large copepods from the training set that were readily confused but not very abundant (e.g. *Eucalanus* sp.), adding larger "discard" particles to the training set, or size fractioning the sample to analyze large particles separately. Appropriate staining of samples prior to analysis might further aid the discrimination of biological from non-biological particles, by raising the importance of optical density in the classification algorithms, and this possibility warrants consideration in future studies.

Small copepods confused the automatic recognition algorithm, possibly due to a size threshold, below which the scanner does not capture sufficient detail (in terms of number of pixels); this was particularly true for small complex particles, i.e. particles with appendages that create irregular outlines. Particles with complex morphology may be more difficult to accurately identify (Culverhouse *et al*., 2003). This may explain the high machine error when classifying small copepods, such as *Pseudocalanus* sp. (copepodite I, II), *Metridia* sp. (copepodite I, II) and *Oithona* sp., as discards (e.g. debris and marine snow) or other small biological particles (e.g. euphausiid eggs, copepod nauplii, small barnacle nauplii).

Finally, euphausiid eggs dominated abundance in some samples and were second only to copepods in many samples, yet ZooImage struggled with their correct identification. Typically, a euphausiid egg appears as a black spherical embryo surrounded by a

clear egg capsule, which may be detected as two particles: one small circular object and one long thin object. Generally, the first was misidentified as phytoplankton and the second as a fiber, both in categories that were removed prior to analysis.

A second goal of this study was to examine how well ZooImage captured the trends of zooplankton taxa and abundance in zooplankton samples. ZooImage was successful at finding patterns in total particles ($R^2 = 0.8779$) and was even better when examining just copepods ($R^2 = 0.9137$) suggesting ZooImage may be capable of capturing trends present in field samples. Furthermore, the relationships described by the regressions indicate no systematic error (i.e. slope = 1), signifying errors in processing were due primarily to machine and user errors. Inconsistency between ZooImage and manual abundances may have several causes. Although the fraction of the sample examined by both approaches was the same, each represented an independent sub-sampling of the collection, so there is not an exact correspondence of particles. Additionally, discards make up the majority of particles in a sample, and since they were deleted any misidentification of organisms as discards (DFP = 1.0%) were also removed, thus lowering the total zooplankton abundance seen by ZooImage. For example, small barnacle nauplii were prevalent in samples from the beginning of the series and were not identified accurately by ZooImage. Ultimately, they were classified as calanoid nauplii and removed from total counts, thus, creating a divergence from manual counts.

This study utilized ZooImage (version 1.2-1) without modification. That is, there was no reconfiguration of the original software to obtain better results. With this configuration, ZooImage requires little technical expertise; the software is designed to lead the user through importing and processing images as well as creating series of ecological parameters such as abundance and biomass. The creation of training sets, both the most time consuming and the most important aspect of ZooImage, requires taxonomic experience to correctly identify organisms placed in the training set. Evaluation through k-fold cross validation and confusion matrices guides the user in creating an effective training set. The time to create a training set will vary depending on the number of different organisms in the environment being studied, the questions being asked and the taxonomic experience of the user.

Preparing the sample for scanning was completed in only minutes and typically consisted of moving particles away from the edges of the image and superficially separating touching particles. ZooImage allows the user to correct for biases in edge particles, but does not correct for user errors. The average user error for bad images or touching particles of 2.33% for all particles (5.40% for biological particles) supports the idea that this bias is minimal compared to machine identification errors; therefore, in practice, little time needs to be spent preparing the samples. Furthermore, the errors generated by misidentification of touching particles are relatively small and did not exceed errors that typically occur when working with zooplankton samples (e.g. splitting errors with expected coefficients of variances between 4.8 and 30.5%) (Guelpen *et al.*, 1992; Postel *et al.*, 2000), and may be offset by the larger number of biological particles that can be enumerated by ZooImage compared to traditional processing.

In summary, the amount of time ZooImage requires to arrive at ecological parameters is much less than the time needed to manually process zooplankton samples. It took <20 min to split and scan each sample, $\sim$35 min for ZooImage to process an entire fractionated sample and only minutes to report abundance (m^{-3}) and biomass. The ability to obtain quantitative results from a zooplankton sample in under an hour demonstrates that ZooImage decreases processing time compared to traditional microscopic analysis which can take several hours. Without demonstrating ability to accelerate sample processing, automated analysis would be of limited utility (Estep and MacIntyre, 1989; Grosjean *et al.*, 2004).

The purpose of this project was to determine the effectiveness of ZooImage as a tool for assessing community composition in preserved zooplankton samples. Processing samples was rapid and efficient and results demonstrated ZooImage's ability to provide useful information. Future versions of ZooImage may incorporate human identification of particles that the software is uncertain about, thus, increasing the accuracy of particle identification. Additionally, because ZooImage is open source software, it can be modified to accommodate many different needs which will ultimately improve the system for all users. At this time, ZooImage shows real promise as a tool for rapid processing of preserved zooplankton samples.

FUNDING

This project was funded by Alaska SeaGrant.

ACKNOWLEDGEMENTS

We wish to thank Dr Philippe Grosjean and Kevin Denis for assistance in using the ZooImage software

and Chris Stark and Cheryl Clark for providing expertise in zooplankton identification. Prince William Sound Aquaculture Corporation supplied the zooplankton samples used in this research.

APPENDIX 1

VueScan setup parameters

[VueScan]	[Color-Perfection4990-Transparency8×10]
[Input-Perfection4990]	WhitePoint=0
Mode=2	WhitePointRed=741
[Input-Perfection4990-Transparency8×10]	WhitePointGreen = 740
Quality=4	WhitePointBlue=744
BitsPerPixel=3	Brightness=1050
MakeGrayFrom=2	[Color]
BatchScan = 1	PixelColors=1
FrameNumber=6	Options=2
PreviewResolution = 7	[Output-Perfection4990-Transparency8×10]
ScanResolution = 3	PrintedSize=0
Mirror=1	[Output]
[Input]	TIFFFile=1
Options=2	TIFFFileName=0061+.tif
[Crop-Perfection4990-Transparency8×10]	TIFFFileType=2
CropUnits=0	TIFFCompression=0
AutoOffset=0	TIFFProfile=0
XOffset=493	JPEGFile=0
YOffset=603	DefaultFolder=C:\\AFK2004
AutoRotate=0	Copyright=UAF/ZooImage2007
CropSize=0	LogFile=0
XSize=16 935	Options=2
YSize=15 120	[Prefs]
XImages=2	GraphType=3
YImages=3	ExternalViewer=0
PreviewArea=0	ExternalEditor=0
PreviewXOffset=2500	WindowXOffset=479
PreviewYOffset=1449	WindowYOffset=216
PreviewXSize=35 000	WindowXSize=1355
PreviewYSize=46 555	WindowYSize=1010
[Crop]	GuidedMode=0
Options=2	StartupTip=0
[Filter]	PreviewMem=471
Options=2	ScanMem=471
	Options=2

REFERENCES

Alvarez-Borrego, J. and Castro-Longoria, E. (2003) Discrimination between *Acartia* (Copepoda: Calanoida) species using the diffraction pattern in a position, rotation invariant digital correlation. *J. Plankton Res.*, **25**, 229–233.

Armstrong, J. L., Boldt, J. L., Cross, A. D. *et al.* (2005) Distribution, size, and interannual, seasonal and diel food habits of northern Gulf of Alaska juvenile pink salmon, *Onchorynchus gorbuscha*. *Deep-Sea Res. II*, **52**, 247–265.

Benfield, M., Davis, C., Wiebe, P. *et al.* (1996) Video Plankton Recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of Georges Ban with comparative measurements from a MOCNESS sampler. *Deep-Sea Res. II*, **43**, 1925–1945.

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

Boldt, J. L. and Haldorson, L. J. (2002) A bioenergetics approach to estimating consumption of zooplankton by juvenile pink salmon in Prince William Sound, Alaska. *Alaska Fish. Res. Bull.*, **9**, 111–127.

Boldt, J. L. and Haldorson, L. J. (2003) Seasonal and geographic variation in juvenile pink salmon diets in the northern Gulf of Alaska and Prince William Sound. *Trans. Am. Fish. Soc.*, **132**, 1035–1052.

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

Castro-Longoria, E., Alvarez-Borrego, J. and Pech-Pacheco, J. (2001) Identification of species of Calanoid copepods using a new invariant correlation algorithm. *Crustaceana*, **74**, 1029–1039.

Cooney, R., Allen, J., Bishop, M. *et al.* (2001) Ecosystem controls of juvenile pink salmon (*Onchorynchus gorbuscha*) and Pacific herring (*Clupea pallasi*) populations in Prince William Sound, Alaska. *Fish. Oceanogr.*, **10**, 1–13.

Culverhouse, P., Ellis, R., Simpson, R. *et al.* (1994) Automatic categorisation of five species of *Cymatocylis* (Protozoa, Tintinnida) by artificial neural network. *Mar. Ecol. Prog. Ser.*, **107**, 273–280.

Culverhouse, 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.

Culverhouse, P., Williams, R., Benfield, M. *et al.* (2006) Automated image analysis of plankton: future perspectives. *Mar. Ecol. Prog. Ser.*, **312**, 297–309.

Davis, C., Gallager, S., Marra, M. *et al.* (1996) Rapid visualization of plankton abundance and taxonomic composition using the video plankton recorder. *Deep-Sea Res. II*, **43**, 1947–1970.

Ellis, R., Simpson, R., Culverhouse, P. *et al.* (1994) Expert visual classification and neural networks: can general solutions be found? Oceans 94. Oceans Engineering for Today's Technology and Tomorrow's Preservation—Proceedings, 1, I.330–I.334.

Embleton, K. V., Gibson, C. E. and Heaney, S. I. (2003) Automated counting of phytoplankton by pattern recognition: a comparison with a manual counting method. *J. Plankton Res.*, **25**, 669–681.

Estep, K. and MacIntyre, F. (1989) Counting, sizing, and identification of algae using image analysis. *Sarsia*, **74**, 261–268.

Estep, K., MacIntyre, F., Hjorleifsson, 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.

Gorsky, G., Guilbert, P. and Valenta, E. (1989) The Autonomous Image Analyzer—enumeration, measurement and identification of marine phytoplankton. *Mar. Ecol. Prog. Ser.*, **58**, 133–142.

Grosjean, P. and Denis, K. (2007) *ZooImage Users Manual*. www//: sciviews.org/Zoo/PhytoImage, last accessed on April 21, 2008.

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.

Guelpen, L. V., Markle, D. and Duggan, D. (1982) An evaluation of accuracy, precision, and speed of several zooplankton subsampling techniques. *J. Cons. Int. Explor. Mer.*, **40**, 226–236.

Hastie, T., Tibshirani, R. and Buja, A. (1994) Flexible discriminant analysis by optimal scoring. *J. Amer. Stat. Assoc.*, **89**, 1255–1270.

Hense, B., Gais, P., Jutting, U. *et al*. (2008) Use of fluorescence information for automated phytoplankton investigation by image analysis. *J. Plankton Res.*, **30**, 587–606.

Hu, Q. and Davis, C. (2005) Automatic plankton image recognition with co-occurrence matrices and Support Vector Machine. *Mar. Ecol. Prog. Ser.*, **295**, 21–31.

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.

Ishii, T., Adachi, M., Omori, M. *et al*. (1987) The identification, counting, and measurement of phytoplankton by an image-processing system. *J. Cons. Int. Explor. Mer.*, 253–260.

Jeffries, H. P., Berman, M. S., Poularikas, A. D. *et al*. (1984) Automated sizing, counting and identification of zooplankton by pattern recognition. *Mar. Biol.*, **78**, 329–334.

Luo, T., Kramer, K., Goldgof, D. *et al*. (2003) Learning to recognize plankton. Proceedings IEEE International Conference on Systems, Man and Cybernetics, Washington, DC, 5–8 October 2003. pp. 888–893.

Luo, T., Goldgof, D., Hall, L. *et al*. (2004) Recognizing plankton images from the Shadow Image Particle Profiling Evaluation Recorder. *IEEE Trans. Syst. Man Cy. B*, **34**, 1753–1762.

Luo, T., Kramer, K., Goldgof, D. B. *et al*. (2005) Active learning to recognize multiple types of plankton. *J. Mach. Learn.*, **6**, 589–613.

Marine Zooplankton Colloquium 2 (2001) Future marine zooplankton research—a perspective. *Mar. Ecol. Prog. Ser.*, **222**, 297–308.

Peters, A., Hothorn, T. and Lausen, B. (2002) Ipred: improved predictors. *R News*, **2**, 33–36.

Postel, L., Fock, H. and Hagen, W. (2000) Biomass and abundance. In Harris, R., Wiebe, P., Lenz, J., Skjoldol, H. and Huntley, M. (eds), *ICES Zooplankton Methodology Manual*. Academic Press, San Diego, pp. 83–174.

Richardson, A. (2008) In hot water: zooplankton and climate change. *ICES J. Mar. Sci.*, **65**, 279–295.

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

Sameota, D., Wiebe, P. H., Runge, J. *et al*. (2000) Collecting zooplankton. In Harris, R., Wiebe, P., Lenz, J. *et al*. (eds), *Zooplankton Methodology Manual*. Academic Press, San Diego, pp. 55–78.

Samson, S., Hopkins, T., Remson, A. *et al*. (2001) A system for high-resolution zooplankton imaging. *IEEE J. Oceanic Eng*, **26**, 671–676.

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

Tang, X., Stewart, W. K., Vincent, L. *et al*. (1998) Automatic plankton image recognition. *Artif. Intell. Rev.*, **12**, 177–199.

Wiebe, P. and Benfield, M. (2003) From the Hensen net toward four dimensional biological oceanography. *Prog Oceanogr.*, **56**, 7–136.

Zavala-Hamz, V., Alvarez-Borrego, J. and Trujillo-Ortiz, A. (1996) Diffraction patterns as a tool to recognize copepods. *J. Plankton Res.*, **18**, 1471–1484.