Methods in Oceanography 15–16 (2016) 60–77

Contents lists available at ScienceDirect

Methods in Oceanography

journal homepage: www.elsevier.com/locate/mio

Full length article

Imperfect automatic image classification successfully describes plankton distribution patterns

Robin Faillettaz [a], Marc Picheral [a], Jessica Y. Luo [b,c], Cédric Guigand [b], Robert K. Cowen [b,c], Jean-Olivier Irisson [a,*]

[a] Sorbonne Universités, UPMC Univ Paris 06, CNRS, Laboratoire d'Océanographie de Villefranche, 181 Chemin du Lazaret, 06230 Villefranche-sur-Mer, France

[b] Marine Biology and Fisheries, Rosenstiel School of Marine and Atmospheric Science (RSMAS), University of Miami, Miami, FL 33149, USA

[c] Oregon State University, Hatfield Marine Science Center, Newport, OR 97365, USA [1]

HIGHLIGHTS

- We propose a simple filtering method to improve classification precision of images.
- It focuses on classification probabilities and discards low confidence objects.
- It successfully resolved the *in situ* fine scale distribution patterns of plankton.
- It requires very limited manual identification work.
- It is applicable to all machine learning methods.

ARTICLE INFO

Article history:
Received 1 August 2015
Received in revised form
13 April 2016
Accepted 16 April 2016
Available online 9 May 2016

Keywords:
Imaging system
ISIIS
Automatic classification

ABSTRACT

Imaging systems were developed to explore the fine scale distributions of plankton (<10 m), but they generate huge datasets that are still a challenge to handle rapidly and accurately. So far, imaged organisms have been either classified manually or pre-classified by a computer program and later verified by human operators. In this paper, we post-process a computer-generated classification, obtained with the common *ZooProcess* and *PlanktonIdentifier* toolchain developed for the ZooScan, and test whether the same ecological conclusions can be reached with this fully automatic dataset and with a reference, manually sorted, dataset.

* Corresponding author.
E-mail address: irisson@normalesup.org (J.-O. Irisson).

[1] Present address.

http://dx.doi.org/10.1016/j.mio.2016.04.003

Plankton distribution
Machine learning
Big dataset

The Random Forest classifier outputs the probabilities that each object belongs in each class and we discard the objects with uncertain predictions, i.e. under a probability threshold defined based on a 1% error rate in a self-prediction of the learning set. Keeping only well-predicted objects enabled considerable improvements in average precision, 84% for biological groups, at the cost of diminishing recall (by 39% on average). Overall, it increased accuracy by 16%. For most groups, the automatically-predicted distributions were comparable to the reference distributions and resulted in the same size-spectra. Automatically-predicted distributions also resolved ecologically-relevant patterns, such as differences in abundance across a mesoscale front or fine-scale vertical shifts between day and night. This post-processing method is tested on the classification of plankton images through Random Forest here, but is based on basic features shared by all machine learning methods and could thus be used in a broad range of applications.

1. Introduction

From the centimetre to kilometre-scales, hydrodynamics, predator–prey interactions and behaviour strongly structure the patchy distributions of planktonic organisms in pelagic environments (Davis et al., 1992; Pinel-Alloul, 1995; Lough and Broughton, 2007). At mesoscales (10–100 km) and submesoscales (<10 km), plankton distributions are primarily determined by hydrological structures like fronts and eddies (Belkin, 2002; Belkin et al., 2009; Luo et al., 2014). For example, convergent flows at frontal features can increase primary production (Grimes and Finucane, 1991) and mechanically concentrate organisms (Bakun, 2006; Olson et al., 1994). However, the influence of these structures may be counter-balanced by behaviour or other biotic processes. Indeed, at fine scale (<1 km), diel vertical migrations can be a strong driver of plankton distributions (Benoit-Bird and McManus, 2012; Neilson and Perry, 1990). At microscales (<1–10 m), biotic interactions such as competition and predation are likely to generate vertical gradients in the distribution of zooplankton. For example, in Monterey Bay, predator avoidance is thought to vertically separate copepods, phytoplankton thin layers, and gelatinous zooplankton predators (Greer et al., 2013). Off the coast of Massachusetts, interactions between internal waves and foraging drive a temporary overlap between layers of high copepod concentration and ichthyoplankton (Greer et al., 2014).

Historically, zooplankton and ichthyoplankton distributions have been sampled with pumps (Herman et al., 1984) and regular or stratified plankton nets (e.g. regular: WP2, Bongo; e.g. stratified: MOCNESS, BIONESS, MULTINET; Wiebe and Benfield, 2003). However, even depth-stratified nets cannot typically resolve the fine and microscale processes at which biotic interactions occur, because they usually sample (and integrate) over at least 10 m vertically and much more horizontally. While pumps offer finer spatio-temporal resolution, they are often limited to surface layers (<10 m depth — Boucher, 1984; sometimes down to 100 m depth — Herman et al., 1984) and sample much smaller volumes (on average 50–60 L min^{-1} vs. 7500 L min^{-1} for a small plankton net; Wiebe and Benfield, 2003).

In the last two decades, *in situ* imaging systems were developed with the aim of sampling microscale processes in the plankton and accelerating data processing using efficient automatic classification techniques (MacLeod et al., 2010; Wiebe and Benfield, 2003). Several imaging systems have emerged, tackling different ecological questions by targeting different size spectra of organisms. The Video Plankton Recorder (VPR; Benfield et al., 1996) and the Underwater Vision Profiler (UVP; Picheral et al., 2010) sample particles and zooplankton. The Shadow Image Particle Profiling Evaluation Recorder (SIPPER; Samson et al., 2001), the ZOOplankton VISualization imaging system (ZOOVIS;

Bi et al., 2013) and the *In Situ* Ichthyoplankton Imaging System, used for this study (ISIIS; Cowen and Guigand, 2008), target large zooplankton up to several centimetres. ISIIS has been specifically designed to sample fish larvae that are patchy and rare (Cowen et al., 2013). Therefore, it samples larger volumes of water compared to other instruments (ISIIS: from 108 to 168 L s^{-1}; UVP: typically 8 L s^{-1}, up to 20.0 L s^{-1}; SIPPER 9.2 L s^{-1}; ZOOVIS 3.6 L s^{-1}; VPR: 10–17 mL s^{-1}) and has proved to be particularly suited to describe the fine-scale distribution of both ichthyoplankton (Cowen et al., 2013; Greer et al., 2014) and other taxa, including gelatinous zooplankton (Luo et al., 2014; McClatchie et al., 2012). These imaging systems generate large datasets of images. For example, in one hour, ISIIS records over 200 billion pixels (the equivalent of more than 200 GB of greyscale TIFF images), usually yielding several hundred million objects of interest, that have to be identified. Manually processing such big datasets has to be limited to few groups of interest (e.g. Greer et al., 2015, 2014; Luo et al., 2014 and McClatchie et al., 2012) but remains time prohibitive. Developing accurate automatic identification processes for such datasets is still a challenge (Benfield et al., 2007; Cowen et al., 2013; Culverhouse et al., 2006) that needs to be solved in order to fully resolve microscale processes.

Imaging data are typically handled in a three-step process: first, detecting and segmenting relevant objects (or regions of interest) from raw images; then measuring features of each object (such as size, aspect ratio, etc.); and finally using these features to classify the objects into biologically/ecologically relevant groups through machine learning algorithms. Several automatic identification procedures have already been tested on plankton datasets of a few thousand images using various classifiers: Random Forest (e.g. Bell and Hopcroft, 2008), Support Vector Machines (e.g. Hu and Davis, 2005), Bayesian models (Ye et al., 2011) or neural networks (e.g. Davis et al., 2004). Some also combined several classifiers to improve prediction accuracy (Hu and Davis, 2005; Li et al., 2014; Zhao et al., 2010). While the algorithms differ, all of these classifiers have in common the fact that they result in a final score (often a probability) for an object to be in *each* class and attribute the object to the class with the highest score. This predicted class is often the only information that is retained from the classifier. So, while classification is typically viewed as a yes-or-no problem, the real outputs from the classifiers are actually continuous.

In this study, we take the example of the commonly-used image processing and identification toolchain *ZooProcess* and *Plankton Identifier* (*PkID*) (Gorsky et al., 2010). The software was first developed for the *ZooScan* (laboratory plankton scanner) and then extended to the UVP (Picheral et al., 2010) and other imaging systems. *ZooProcess* segments objects from the full image and computes a set of descriptive features (grey levels, length, width, area, shape, etc.) that are then used by *PkID* through various classification algorithms (Support Vector Machine, Neural network, Random Forest, etc.), although Random Forest (Breiman, 2001) has proven to be the most accurate and is now used routinely (Gorsky et al., 2010). This software suite is free, open-source, easy to install, and well supported. Therefore, it is widely distributed worldwide and used by 60 research teams from the tropics to the poles (e.g. France (Vandromme et al., 2011); New-Caledonia (Smeti et al., 2015) Antarctica (Espinasse et al., 2012)). It is most commonly used as a *computer-assisted* identification system, whereby the classifier proposes identifications that are then validated by human operators for all objects.

ZooProcess and *PkID* offer appropriate tools to handle ISIIS data but the amount of data generated by ISIIS makes human validation impractical. For example, validating the identifications of the 1.5 million objects used as a reference in this study took seven full-time months; a few days of ISIIS deployments typically yield from ten to a hundred million objects. However, given the size and spatial resolution of the dataset, even a subset of it is likely to contain relevant ecological information, at least at the metre to 10 m scale. Here, we propose to discard objects with a low classification score (i.e. the least likely to be correctly identified) and assume that all remaining objects are correctly classified, hence bypassing the validation step. Most other studies compare automatic classification methods using only classification metrics (e.g. precision, recall). We suggest that a more biologically relevant approach is to examine whether the same ecological patterns can be detected in datasets generated by various methods. Here we compare the same data either manually identified (hereafter the *reference* dataset) or automatically classified and further filtered based on classification score (hereafter the *predicted* dataset). We specifically explore the fine-scale spatial distribution of zooplankton across a frontal structure, its relationship with the environment, the size distribution of planktonic groups as well as their diel vertical migration patterns.

2. Materials and methods

2.1. Description of ISIIS

The *In Situ* Ichthyoplankton Imaging System (ISIIS) is a towed underwater imaging system (Cowen and Guigand, 2008). It uses backlight shadowgraph imaging, which makes it ideally suited for small and often transparent planktonic organisms in a consistent manner. The version of ISIIS used here was slightly modified from that of Cowen and Guigand (2008). The line-scan camera imaged a 10.5 cm-tall field of view with a 50 cm depth of field. With a line-scan camera, the image is created by the movement of the instrument and scanning at 28 kHz produced a continuous image when towed at 2 m s^{-1} (4 knots). These settings resulted in a sampling rate of 108 L s^{-1}. Additionally, ISIIS is equipped with environmental sensors recording temperature, conductivity (hence salinity and density), oxygen, chlorophyll *a* fluorescence and photosynthetically active radiation (PAR) at a rate of 2 Hz.

2.2. Test data

ISIIS was deployed for two transects across the Ligurian current, a coastal jet that creates a permanent, mesoscale front. The current delineates a coastal, a frontal and an offshore zone, with characteristic hydrological properties (Sammari et al., 1995) and biological communities (Boucher et al., 1987). One transect was conducted at night, the other during the following day, in July 2013. Both transects were conducted on the same line, though the night transect sampled from onshore to offshore, and the day transect sampled from offshore to onshore. Thanks to moveable fins, ISIIS sampled the water column in a tow-yo fashion, between the surface and 100 m depth, with a vertical speed of 0.2 m s^{-1}. The images in this study come from 13 down-casts of the night transect and 7 down-casts of the day transect, which were the only ones fully processed of the ~26 total up- and down- casts of each transect.

2.3. Image pre-processing

ISIIS collected a continuous stream of pixels, 2048 pixels in height. The stream was cut into square 2048 × 2048 frames by the acquisition software (example in Fig. 1). Because the camera was continuously scanning the same line, a single speckle or scratch along the optical path would create a continuous streak in the resulting 2D image. These streaks were removed by dividing each frame by the average of the previous 50 consecutive frames and normalising the result to [0, 255] in grey intensity, a process known as flat-fielding.

2.4. Segmentation

The shadows of planktonic organisms or particles imaged by ISIIS appeared dark on a light background. All images were thresholded at the 195 grey level; i.e. adjacent pixels darker than 195 (255 = white, 0 = black) were considered as objects of interest. The flat-fielding procedure resulted in an almost white background and well contrasted objects (Fig. 1). Therefore, the detection of objects was not very sensitive to the threshold value and 195 was chosen after a few tests.

Small objects were difficult to identify reliably, even for human operators. Only objects larger than 250 px in area (equivalent to 18 px in diameter for a spherical object) were considered in this study. With a pixel resolution of 51 μm, this converts to an area of 0.6 mm^2 and an equivalent diameter of 920 μm.

All objects with sufficient size and darkness were segmented out of the frames (Fig. 1 exemplifies which objects were considered and which were not) and the region outside of the object itself was made pure white. A total of 1.5 million objects were detected.

2.5. Feature extraction

The purpose of this study is to optimise an existing classification procedure *a posteriori*. Therefore, the feature extraction was based on the standard configuration in *ZooProcess/PkID* and is not described

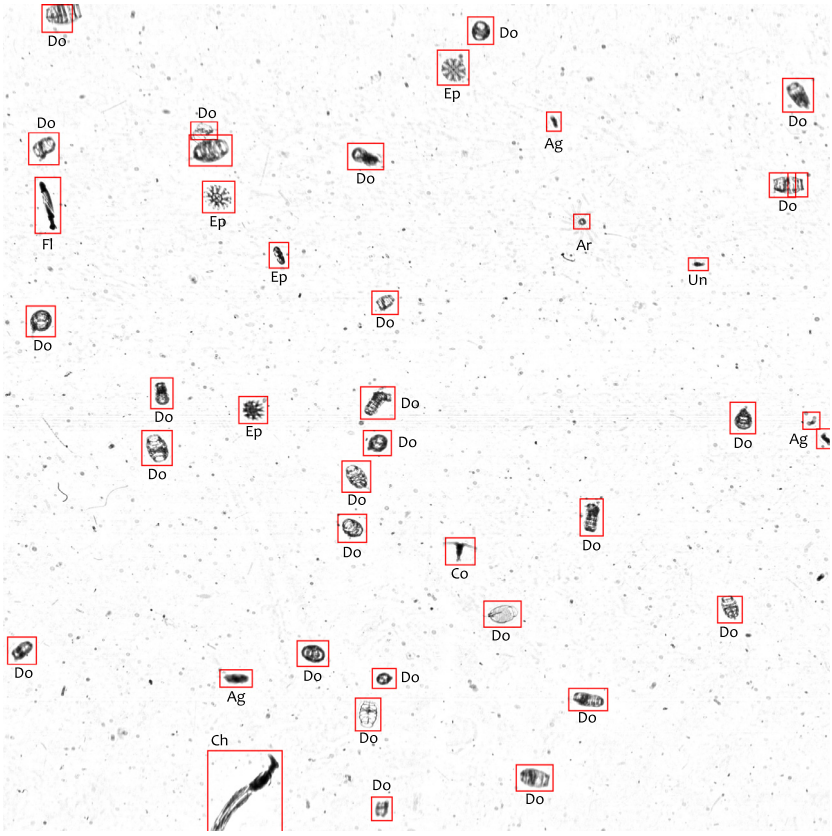

Fig. 1. Example of a flat-fielded 2048 × 2048 pixels frame collected by ISIIS. The bounding box of objects extracted and measured is drawn. Those objects are labelled (Ag: aggregates; Ar: Trachymedusae *Arctapodema spp*; Ch: chaetognath; Co: calanoid copepod; Do: doliolid; Ep: *Pelagia noctiluca* ephyrae; Fl: fish larva; Un: unidentified). Note that, on rare occasions, some small-bodied and transparent organisms, such as doliolids, were either truncated or split into several objects and then became hardly identifiable.

in detail here (please refer to Gasparini and Antajan, 2013 and Gorsky et al., 2010). Briefly, 37 features were measured by *ZooProcess,* and 9 additional variables were derived by PkID from the original 37 features. These features characterised each object's size and shape (length of the minor and major axes of the best fitting ellipse, Feret diameter, circularity, symmetry, aspect ratio), transparency (five measures of grey levels: mean, mode, standard deviation, minimum, maximum), and aspect (grey level histogram descriptors such as skewness, cumulative histograms, etc.). When combined, those features can characterise object classes; for example, small, dark, ovoid objects with a large Feret diameter compared to their overall size are probably copepods with their antennae extended. Therefore, they serve as the basis for automatic classification.

2.6. Learning set and classification

Supervised classification techniques require a set of identified and measured objects to learn the differences between classes based on their features. Our learning set comprised 14 biotic and abiotic classes with a target size of 200 objects per class (see Table 1), a number which proved to be appropriate for previous *ZooProcess/PkID* projects (Gorsky et al., 2010). The most numerous classes in the data (noise in particular) were also inflated in the learning set, to get a total of 5979 objects. Objects in the learning set were chosen to be representative of the diversity of each class.

Table 1
Name, number of objects in the learning set (n) and description of classes. First non-living objects or artefacts, then biological organisms.

Class	n	Description, taxonomical identification
Dark aggregates	314	Solid, opaque marine snow
Light aggregates	489	Marine snow (larvacean houses, mucus, etc.)
Fibres	433	Thin fibres and faecal pellets
Noise	2296	Noise generated by water density changes
Tentacles	224	Pelagia noctiluca tentacles
Copepods	349	Mainly calanoid copepods
Doliolids	209	Thaliacean, Family Doliolidae
Fish larvae	289	Fish larvae
Trachymedusae	200	Trachymedusae (e.g. *Arctapodema spp*)
Diatom chains	342	Phytoplankton, diatoms chains
Acantharian radiolarians	213	Radiolaria, Order Acantharia
Radiolarian colonies	255	Radiolaria, Order Colodaria, incolonies
Solitary radiolarians	267	Radiolaria, Order Colodaria, solitary
Shrimps	99	Shrimp-like organisms (e.g. Mysidacae or Euphausiacae)

All 1.5 million segmented objects were classified into these 14 classes by a Random Forest classifier using the 46 measured features (Gorsky et al., 2010). The parameters of the classifier were left at the appropriate defaults in *PkID*: 100 trees, bagging of 1, 6 features randomly selected per tree, leaf size of 2 objects.

Finally, three trained operators validated the classification of each object, yielding a completely manually-identified dataset of 1.5 million objects, hereafter referred to as the *reference* dataset.

2.7. Data filtering and optimisation of the classifier precision

To detect meaningful ecological patterns in the distribution of a computer-predicted class, there needs to be sufficiently high confidence that objects in that class belong to the same taxonomic group. In terms of classifier performance, this requires high *precision* (precision = proportion of correctly classified objects in a predicted class). With low precision, a predicted class would be a heterogeneous mixture of various taxonomic groups, the distribution of which cannot be interpreted ecologically. Conversely, for high frequency imaging datasets, the data are often in sufficient quantity that a subsample of the whole dataset would be enough for detecting ecological patterns. In terms of classification metrics, a low *recall* may be acceptable (recall = proportion of the total number of objects of a class that are predicted in that class). Therefore, we suggest that, to detect ecological patterns in a high frequency dataset, particularly for common taxa, precision is more important than recall. To test this hypothesis, we filtered out the most likely mistakes in the computer-predicted dataset (to increase precision), at the cost of discarding some correctly identified objects (hence decreasing recall), and then compared the resulting dataset against the reference set.

The probabilities for each object to be in each class (i.e. the final output of the classifier) were used as the filtering criterion. All objects assigned to a given class were ranked in increasing order of probability. All objects with probability above a threshold were kept and assumed to be correctly identified; other objects, with probability equal to or lower than the threshold, were considered to be potentially wrong and were discarded. Since precision needs to be controlled, the threshold should be set to result in a given precision. For example, picking the probability of the first wrongly identified object as the threshold would yield 100% precision (all objects ranked above the first false positive are correctly classified). Here, a 1% error rate (99% precision) was deemed acceptable. Error rates lower than 1% resulted in discarding 3% more objects while improving precision by only 0.2. Higher error thresholds resulted in low precision when applied to the whole dataset (average precision with threshold at 10% = 54, at 5% = 60.1, at 1% = 76.9). A 1% error threshold allowed us to increase precision significantly and still keep a representative percentage of objects.

The computation of thresholds was done with the learning set only, because in operational conditions, only the identifications of the objects in the learning set are known. The class probability

of each object in the learning set was predicted using 2-fold cross-validation repeated 50 times, using the Random Forest classifier in PkID. The probabilities were averaged over the 50 repetitions, objects were assigned to the class of highest probability, and probability thresholds at 1% error were computed in each class. Those thresholds, computed on the learning set, were then applied to the predictions of the 1.5 million objects and the subset of objects that was kept constituted the *predicted* dataset. Thus, once the objects in the learning set are identified manually (which is required for prediction anyway), this precision optimisation method requires only computation, no further human validation effort.

2.8. Consequence of data filtering on classification metrics

By construction, the chosen thresholds resulted in exactly 99% precision on the learning set. Because all 1.5 million objects in the reference set were actually identified in this exercise, the precision, recall and F1 score ($2 \times \text{precision} \times \text{recall}/(\text{precision} + \text{recall})$) could be computed for each class over the whole dataset, before and after the filtering process. This allowed us to check whether the precision after filtering approached 99% on the whole dataset as well and how much this improvement in precision costs in terms of decrease in recall.

2.9. Comparison of size spectra

The size structure of planktonic communities is often considered as a proxy to study the transfer of energy through the food web and the export and sequestration of carbon (Legendre and Le Fèvre, 1991). It could be expected that smaller objects would be less defined, would therefore be predicted with lower confidence (i.e., lower probabilities) and may be preferentially filtered out by our method. To assess this, size spectra (i.e., probability density distributions of sizes) were estimated with a kernel method (Gaussian kernel with a 0.25 mm standard deviation) and compared in the reference and predicted dataset.

2.10. Statistical comparisons of spatial distributions

Individual objects were counted over 1 m depth bins along the undulating trajectory of ISIIS and counts were transformed into concentrations by dividing by the volume sampled in each bin. This resulted in maps of the concentration of each class of organism across depth (0–100 m) and distance from the coast (0–60 km) for each transect (for examples see Figs. 3 and 4).

The similarity between the maps for the reference and predicted datasets was assessed using the t-test modified by Dutilleul (Dutilleul et al., 1993; H0: no correlation between the maps, H1: significant correlation between the maps), as well as the Pearson and Spearman correlation coefficients. On a map, observations close to each other are usually similar; this spatial autocorrelation means that observations close to each other are not independent and that the number of actual degrees of freedom is lower than the apparent sample size. The Dutilleul t-test corrects the number of degrees of freedom based on the spatial autocorrelation of the data (computed as Moran's I) and is therefore appropriate to avoid over-estimating the similarity of spatial patterns.

Because diel-vertical migration is such a widespread behaviour in marine ecosystems (Hays, 2003) and strongly influences survival through predator-avoidance and foraging in many taxa (Neilson and Perry, 1990), data were specifically inspected in the vertical dimension. Average vertical distributions were computed for each group and each transect (hence separating day and night). Reference and predicted vertical distributions were compared with the version of Kolmogorov–Smirnoff test modified by Solow et al. (2000), which specifically takes into account autocorrelation along depth caused by the patchiness of plankton.

By construction, concentrations were lower in the predicted dataset than in the reference dataset, because the former is a subset of the latter. Before the comparisons described above, concentrations were normalised to a maximum value of 1 for each class in each transect, by dividing by the maximum concentration recorded. This puts the focus on distribution patterns, rather than actual concentration values, which were poorly estimated when recall was low anyway.

Finally, the predicted and reference datasets are not independent (one is a subset of the other) and the absolute values of the test statistics and *p*-values are therefore biased. The relative values, among classes, are informative however.

2.11. Comparison of ecological patterns

The frontal structure across which the transects were sampled is characterised by an inshore–offshore gradient of increasing salinity, with a front that can be delineated by the 38.2 and 38.3 isohalines (Sammari et al., 1995) and is expected to strongly structure zooplankton communities (e.g. Boucher, 1984 and Pedrotti and Fenaux, 1992). Beyond comparing the distribution maps for the reference and predicted datasets statistically, the results were interpreted with respect to the frontal structure to check whether the ecological patterns were the same. In addition, the relationships between planktonic abundances and environmental variables were inspected in the reference and predicted datasets. The variables inspected were: salinity, which best marks the front, temperature, which is strongly stratified vertically, chlorophyll *a* fluorescence, which marks a clear Deep Chlorophyll Maximum (DCM), and oxygen concentration, which depends both on the frontal structure and on the DCM. When the relationships could be considered linear, the slopes were estimated through Generalised Linear Models (GLM) with Poisson errors and statistically compared between the two datasets using ANOVA.

Similarly, beyond comparing vertical distributions statistically, we assessed whether the range and strength of diel vertical migrations could be as readily detected in the predicted dataset than in the reference dataset. Within each class, day and night distributions were compared with the Solow–Kolmogorov–Smirnov test and the value of its statistic was compared between reference and predicted data. The day–night shift in the depth centre of mass of the distributions (mean of depth weighted by abundance at that depth, Z_{cm}; Irisson et al., 2010) was computed and compared between the reference and predicted datasets.

2.12. Data selection

Abrupt changes in water temperature around the thermocline generated large density differences, which are unfortunately well captured on shadowgraphs. These numerous objects ($n = 1{,}287{,}302$) were classified as "Noise". Another abundant class of objects were tentacles of the medusa *Pelagia noctiluca* ($n = 8106$), which occasionally got stuck on ISIIS and were imaged constantly. These two classes of objects are not biologically relevant in the present study, but were abundant and predicted with high precision (>95%), and were thus both omitted from the subsequent analyses.

3. Results

3.1. Consequences of data filtering on classification metrics

Discarding low probability images considerably increased precision, by 37% on average (Table 2). While probability thresholds were set to yield 99% precision on the cross-validated learning set, precision was lower when the thresholds were applied to the whole dataset. This was expected, because the ~6000 images in the learning set cannot fully represent the variability in the whole dataset (1.5 million images). The average precision of the biological categories after filtering was 84%. The trachymedusae and Acantharian radiolarians displayed the lowest precision (61.9% and 65.4% respectively) but this already was an improvement of more than 50% compared to the situation before filtering.

To reach these precision levels, a large amount of images had to be discarded, leaving only 28.1% of the objects from the original dataset ($n = 39{,}758$, excluding "noise" images). The percentage of objects retained ranged from 8.5% for fibres ($n = 557$) to a maximum of 63.7% for solitary radiolarians ($n = 8569$). As a consequence, on average, filtering decreased recall by 39% and F1 score by 7.8%. However, the improvement in precision dominated the effect of the decrease in recall, because classification accuracy of the whole dataset improved from 40.2% to 56.3% after filtering.

Table 2
Classification metrics before and after filtering out objects with low prediction confidence: number of particles before filtering (n); percentage of data kept after filtering; precision, recall, and F1 score before and after filtering, and difference (after–before). Improvements (positive differences) are bolded. Non-living groups are presented first, groups of biological interest second.

Class	n	%kept	Precision			Recall			F1		
			Before	After	Diff	Before	After	Diff	Before	After	Diff
Dark aggregates	60 164	6.5	77	95	**19**	50	7	−43	60	7	−54
Light aggregates	4 209	4.2	8	17	**9**	53	4	−49	14	4	−10
Fibres	8 055	6.9	46	85	**38**	56	7	−49	51	7	−44
Copepods	17 459	22.4	54	88	**34**	72	22	−49	62	22	−39
Doliolids	30 478	40.2	80	95	**16**	64	40	−24	71	40	−31
Fish larvae	802	23.2	12	80	**67**	62	23	−39	21	23	**3**
Trachymedusae	524	50.6	9	62	**53**	79	51	−29	16	51	**35**
Diatom chains	11 015	28.6	75	97	**22**	72	29	−43	73	29	−45
Acantharian radiolarians	1 021	18.9	7	65	**58**	74	19	−55	14	19	**5**
Radiolarian colonies	4 367	16.7	24	94	**70**	62	17	−45	35	17	−18
Solitary radiolarians	13 049	65.7	68	88	**19**	89	66	−23	77	66	−12
Shrimps	213	52.6	51	89	**38**	74	53	−21	60	53	−7

3.2. Comparison of size spectra in the reference and predicted datasets

In most classes, the size distribution of objects in the automatically predicted dataset and in the reference dataset was closely related (Fig. 2). However, in three groups (fish larvae, radiolarian colonies, and shrimps), the shape of the spectrum was conserved but the occurrence of small objects was under-estimated. In particular, the mode of the spectrum (i.e. the most frequent size class) was larger by 1.3 mm for fish larvae in the predicted dataset compared to the reference dataset, by 6 mm for radiolarian colonies and by 2.8 mm for shrimps (Fig. 2).

3.3. Distribution of plankton with respect to the front

The automatically predicted and filtered spatial distributions of most taxa and particles were significantly correlated with the reference distributions in 20 of the 22 groups at the $p < 0.001$ level (Table 3; Fig. 3). Correlation coefficients were also very high (seven classes with $r > 0.7$, and eight additional classes with $r > 0.5$). The only two exceptions are fish larvae and shrimps in the day transect, both of which were very rare.

At the chosen 99%-precision filtering level, so many images of fish larvae and fibres were discarded that the resulting spatial distributions were very sparse (14.9% and 8.5% of images left, respectively; Fig. 4). Such sparse distributions would clearly not be interpreted ecologically, given how little data are left and how much is discarded. So, information is lost but at least no wrong conclusions would be drawn. In addition, even in those cases, the locations of the maximum concentration zones were properly captured in the predicted dataset; there were just too few objects to represent the finer patterns (Fig. 4).

The reference spatial distributions showed that most taxa were strongly influenced by the frontal zone: fish larvae, Acantharian radiolarians and doliolids were constrained on the coastal side of the front, copepods were also more concentrated towards the coast and in the upper layers of the water column, while diatom chains were more abundant in the deep, offshore zones (Fig. 3, left column). The high spatial resolution of the data allowed us to detect smaller scale patterns such as a region of slightly lower concentrations of copepods and solitary radiolarians at the front (around 30 m depth for copepods and 50 m depth for radiolarians; Fig. 3). Solitary radiolarians also occurred in shallower water in the offshore zone compared to the coastal zone (Fig. 3) and precisely followed the DCM (not mapped). All these patterns, from the contrasts between taxa to the fine-scale low concentration regions at the front, could also be well detected on the predicted data (Fig. 3, right column). The ecological interpretations in terms of the distribution relative to the frontal zone would be the same.

The relationships between the abundance of biological taxa and various environmental variables (salinity, temperature, chlorophyll *a* fluorescence, oxygen concentration) were very similar in the

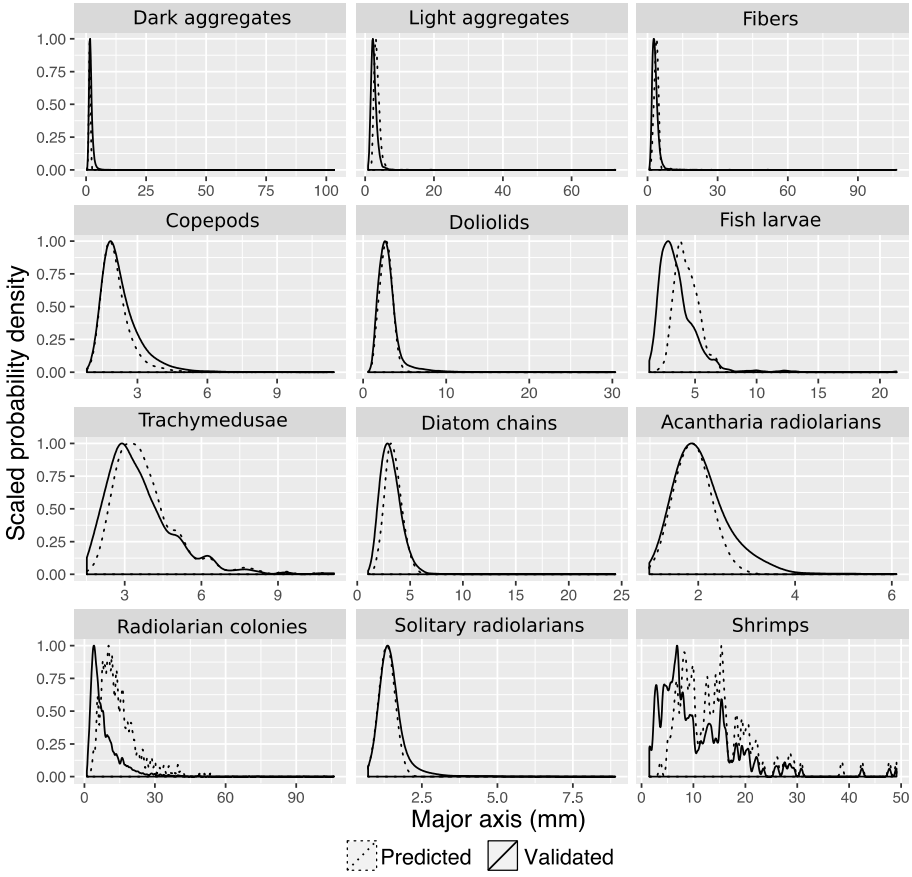

Fig. 2. Per-class size spectra in the reference (solid lines) and automatically predicted and filtered (dotted lines) datasets. Probability density distributions of sizes were scaled between 0 and 1 to focus attention on the shapes of the distribution rather than the differences in the number of objects between the two datasets. The minimum size of objects considered was 250 pixels in area, resulting in $\geq$920 μm in major axis.

reference and predicted datasets. In fact, in 69 of the 80 relationships that could be modelled with GLMs, the slopes were not significantly different between the two datasets. For example, copepods were more abundant in fresher waters (Fig. 5), which were found on the coastal side of the front. The relationships with chlorophyll *a* fluorescence highlighted the association of diatom chains and solitary radiolarians with the DCM. Finally, doliolids were vastly more abundant in warmer, surface waters (Fig. 5). All these conclusions would be reached with the predicted dataset, which suggests that it could be used to explore and define the habitat preference of various organisms.

3.4. Day and night vertical distributions

In 8 of 12 groups, the predicted and reference vertical distributions were slightly but significantly different (Solow–Kolmogorov–Smirnov test, $p < 0.05$; Table 4). The four groups in which the distributions were not statistically different were doliolids, Acantharian radiolarians, colonial radiolarians and shrimps, although the lack of significant difference in the latter group was probably due to their low overall numbers.

For many groups, except trachymedusae and fish larvae, ecological conclusions regarding depth spread and preferendum would be the same in the reference and predicted dataset, even when

Table 3
Statistical comparisons of spatial distributions between the reference and predicted datasets with three statistics: Dutilleul modified t-test (statistic, recomputed degrees of freedom and p-value), Pearson's correlation coefficient and Spearman's rank correlation coefficient. NB: no light aggregates were observed at night.

Class	Transect	Dutilleul t-test			Pearson's r	Spearman's rho
		F-stat	DoF	p-value		
Dark aggregates	Night	29.99	35	$\mathbf{p < 0.001}$	0.66	0.68
	Day	24.11	20	$\mathbf{p < 0.001}$	0.68	0.74
Light aggregates	Day	10.05	76	$\mathbf{p < 0.01}$	0.11	0.34
Fibres	Night	103.22	155	$\mathbf{p < 0.001}$	0.38	0.62
	Day	144.93	191	$\mathbf{p < 0.001}$	0.42	0.62
Copepods	Night	54.37	36	$\mathbf{p < 0.001}$	0.74	0.71
	Day	36.50	28	$\mathbf{p < 0.001}$	0.73	0.71
Doliolids	Night	12 244.11	275	$\mathbf{p < 0.001}$	0.66	0.94
	Day	27 064.77	187	$\mathbf{p < 0.001}$	0.55	0.94
Fish larvae	Night	231.25	162	$\mathbf{p < 0.001}$	0.44	0.77
	Day	1.58	561	0.21	0.09	0.05
Trachymedusae	Night	286.28	168	$\mathbf{p < 0.001}$	0.61	0.78
	Day	130.66	287	$\mathbf{p < 0.001}$	0.48	0.55
Diatom chains	Night	431.64	74	$\mathbf{p < 0.001}$	0.72	0.92
	Day	377.12	97	$\mathbf{p < 0.001}$	0.75	0.86
Acantharian radiolarians	Night	130.32	176	$\mathbf{p < 0.001}$	0.53	0.64
	Day	107.86	167	$\mathbf{p < 0.001}$	0.47	0.65
Radiolarian colonies	Night	220.39	358	$\mathbf{p < 0.001}$	0.61	0.64
	Day	116.20	393	$\mathbf{p < 0.001}$	0.52	0.49
Solitary radiolarians	Night	107.11	22.24	$\mathbf{p < 0.001}$	0.91	0.89
	Day	101.06	14.33	$\mathbf{p < 0.001}$	0.92	0.91
Shrimps	Night	685.26	893.08	$\mathbf{p < 0.001}$	0.72	0.82
	Day	0.01	719.25	0.91	0.00	0.00

Table 4
Statistical comparisons of vertical distributions between the reference and predicted datasets. The statistic and p-value of the Solow–Kolmogorov–Smirnov test are reported, as well as the depth centre of mass of the distribution.

Class	Transect	Solow–K–S			
		Reference–predicted		Depth (m)	
		K	p	Reference	Predicted
Dark aggregates	Day	3.22	$<$**0.0001**	49.1	55.3
	Night	3.91	$<$**0.0001**	41.2	53.1
Light aggregates	Day	2.98	$<$**0.0001**	29.0	40.5
Fibres	Night	3.97	$<$**0.0001**	51.5	69.3
	Day	1.61	**0.0050**	61.8	69.7
Copepods	Night	2.97	$<$**0.0001**	40.8	44.9
	Day	1.44	**0.0250**	56.1	55.1
Doliolids	Night	0.67	0.5690	5.1	6.9
	Day	0.82	0.3370	7.1	8.6
Fish larvae	Night	1.86	$<$**0.0001**	16.9	10.9
	Day	1.25	**0.0490**	32.6	52.2
Trachymedusae	Night	1.44	**0.0080**	10.5	12.7
	Day	1.31	**0.0240**	25.9	29.5
Diatom chains	Night	3.67	$<$**0.0001**	57.5	63.1
	Day	1.72	**0.0010**	64.3	67.8
Acantharian radiolarians	Night	1.13	0.1300	25.3	27.1
	Day	0.69	0.6070	28.3	29.9
Radiolarian colonies	Night	1.20	0.0940	45.4	44.4
	Day	0.51	0.9020	45.8	46.3
Radiolarians solitary	Night	2.43	$<$**0.0001**	53.5	55.9
	Day	2.23	$<$**0.0001**	59.3	60.9
Shrimps	Night	1.00	0.1990	55.3	53.8
	Day	0.51	1.0000	49.9	44.1

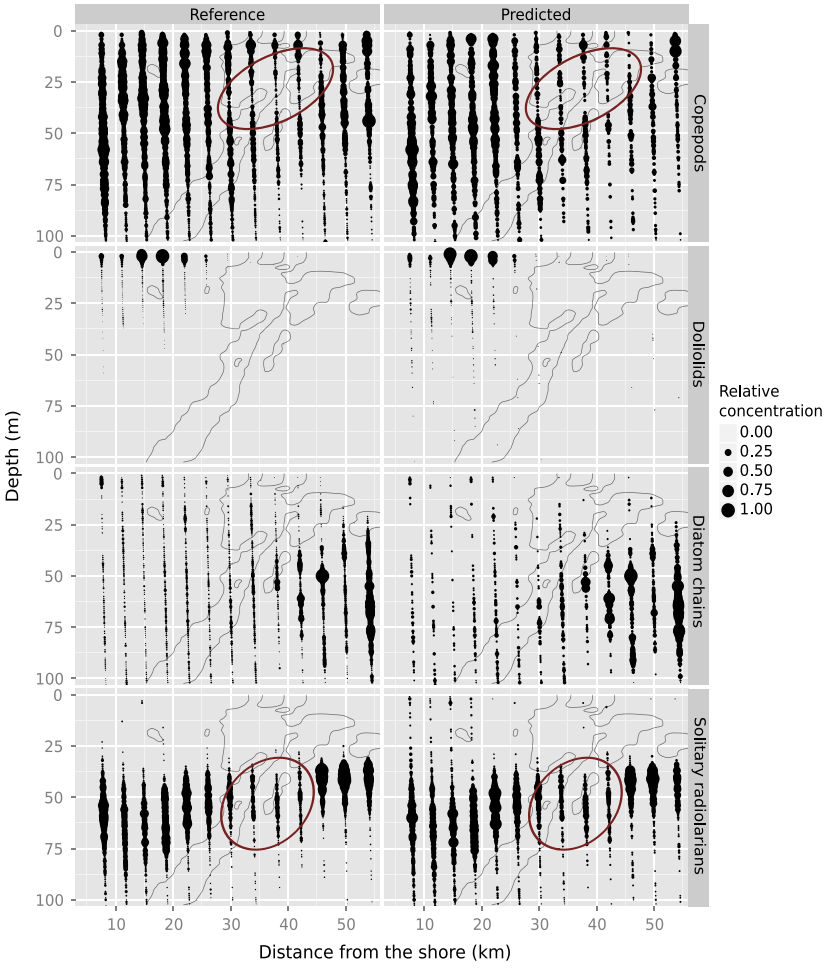

Fig. 3. Examples of some spatial distributions in the predicted dataset (right) that are well correlated with the reference dataset (left). From top to bottom: copepods, doliolids, diatom chains and solitary radiolarians, all during the night transect. The *x*-axis is the distance from the coast (coastal side on the left, offshore side on the right). The area of the dots is proportional to the concentration, scaled to a maximum of 1 per taxon in each dataset, to ease comparison of patterns; the legend shows five examples but scaling is continuous. Grey lines are the 38.2 and 38.3 isohalines that delineate the frontal region. Ellipses highlight regions of lower concentration located in the frontal zone.

distributions were statistically different (Table 4, column "Depth (m)" and Fig. 4). Similarly, an analysis of diel vertical migration patterns would reach very similar conclusions on the reference and on the predicted dataset. When a significant diel vertical migration was detected in the reference dataset, it was also significant in the predicted one (Table 5). Conversely, radiolarian colonies and Acantharian radiolarians do not appear to vertically migrate and this conclusion was also reached with the predicted dataset. The range of downward migration of Trachymedusae, solitary radiolarians and doliolids was also very comparable between the datasets; the same was true, to a lesser extent, for calanoid copepods (Table 5, Fig. 6). However, the vertical migration of fish larvae was poorly predicted, with a bias towards the surface at night that was much greater than in reality (Fig. 6).

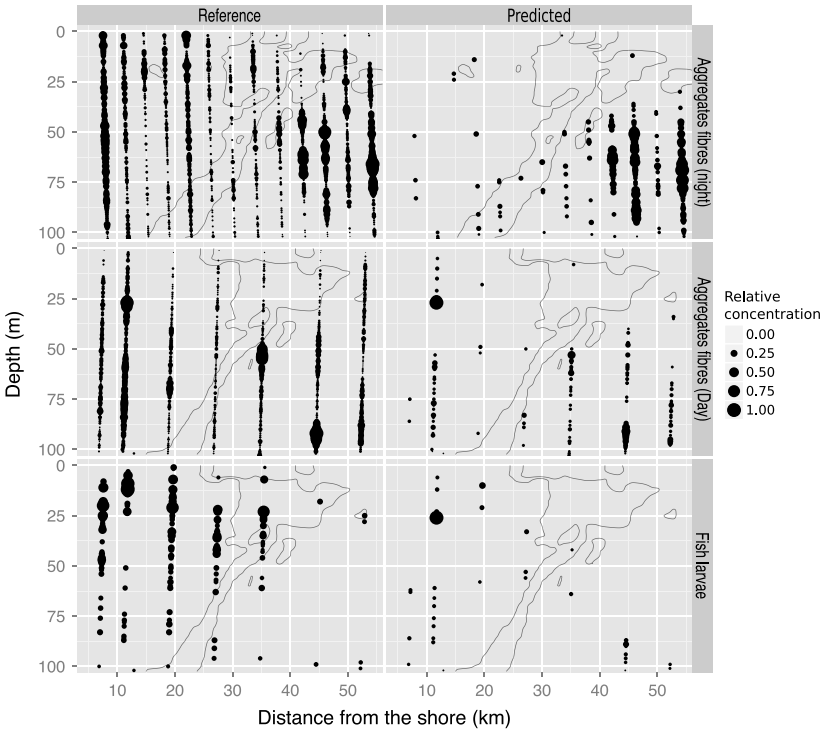

Fig. 4. Examples of poorly predicted spatial distributions (right) compared to the reference distributions (left). From top to bottom: fibres at night, then during the day and fish larvae during the day. Same conventions as Fig. 3.

Table 5
Comparison of the resolution of diel vertical migration patterns in the reference and predicted datasets. Reported for each dataset are: (i) the statistic (K) of the Solow–Kolmogorov–Smirnov test comparing day and night (bold when the test is significant), which quantifies the overall difference in distribution, and (ii) the difference between the depth centre of mass at night and during the day, a proxy for the migration range (night–day; negative means upward migration at night).

	Solow–K–S day–night (K)		Migration range (m)	
	Reference	Predicted	Reference	Predicted
Copepods	**4.10**	**2.86**	−15.3	−10.3
Doliolids	**1.16**	**1.14**	−2.1	−1.7
Fish larvae	**1.88**	**1.72**	−15.8	−41.4
Trachymedusae	**1.72**	**2.07**	−15.4	−16.8
Diatom chains	**2.53**	**2.25**	−6.8	−4.7
Acantharian radiolarians	0.99	1.15	−3.0	−2.9
Radiolarian colonies	0.50	0.67	−0.4	−1.9
Solitary radiolarians	**3.04**	**2.75**	−5.8	−5.0
Shrimps	0.83	0.81	5.4	9.6

4. Discussion

The method presented here aimed at bypassing the manual validation of predicted identifications by discarding objects classified with low confidence, hence improving precision (but decreasing recall). The precision increase (+37% on average) was counter-balanced by a recall decrease (−39% on average), but overall classification accuracy using this method increased by 16%.

The quality and resolution of images may influence the maximum taxonomic resolution achievable by any automatic classification method. Studies based on high quality laboratory imagery of plankton

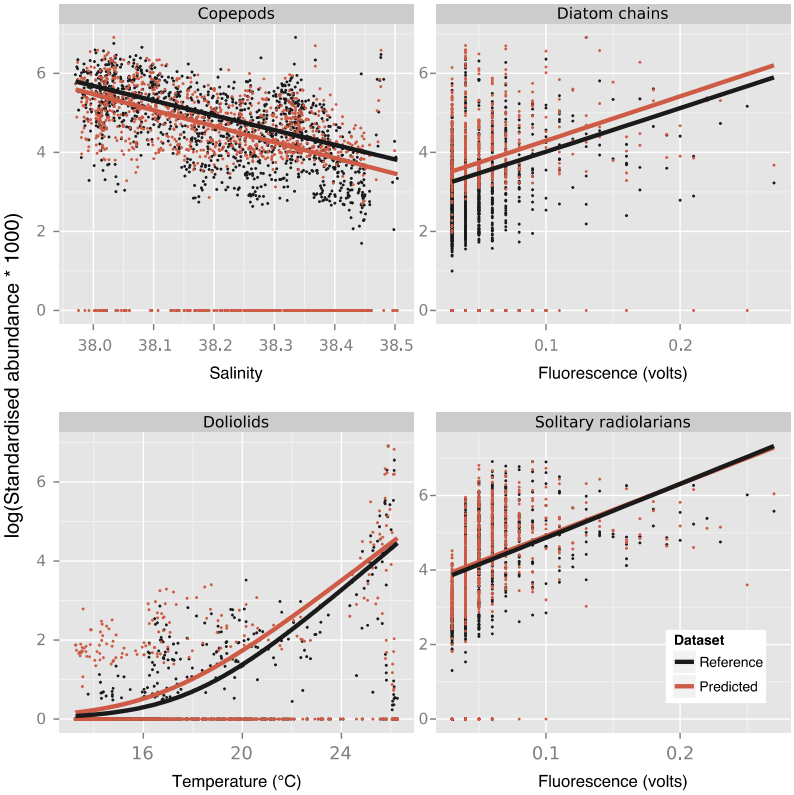

Fig. 5. Examples of the influence of environmental variables on the distribution and concentration of several taxa for the reference dataset (black) and automatically predicted and filtered dataset (red). The lines are the fitted values of GLMs with a Poisson distribution of the residuals. The slopes of the GLM based on the predicted dataset are not significantly different from the ones based on the reference dataset (ANOVA, all $p > 0.05$). Concentration is standardised between groups based on the maximum concentration per taxa and per dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

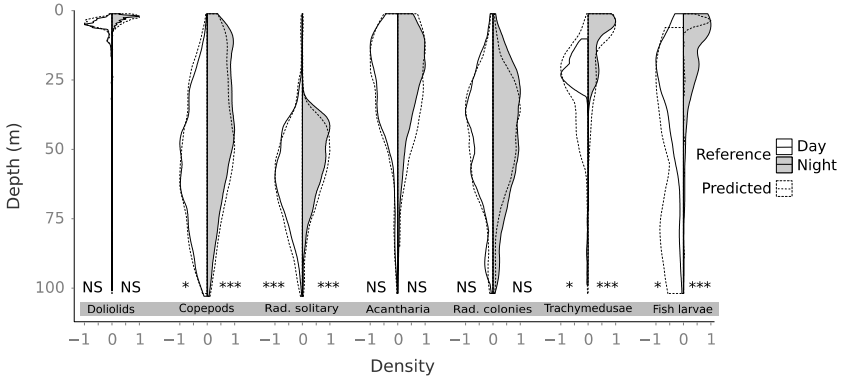

Fig. 6. Examples of vertical distribution during the day (left side) and at night (right side, shaded) as depicted in the reference dataset (solid) and in the predicted and filtered dataset (dashed). The significant levels of the comparisons between reference and predicted distributions are indicated for both day and night (NS: not significant; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$).

have usually reached higher accuracy and could resolve a larger number of groups (e.g. 22 phytoplankton groups in Sosik and Olson, 2007; 25 zooplankton groups in Fernandes et al., 2009; 10–20 groups in Benfield et al., 2007) than studies based on images of zooplankton captured *in situ* which are usually of lesser quality (e.g. three groups with SVM, achieving 80% accuracy (Bi et al., 2015); seven groups with random subspace model achieving >90% precision but in a self-prediction of the learning set (Zhao et al., 2010); five to seven groups with neural networks, reaching 60%–80% accuracy; Davis et al. (2004) and Hu and Davis (2005)). While only a formal comparison, using the same dataset (e.g. Fei-Fei et al., 2007), could resolve the differences between classification methods, comparing the size orders of classification metrics between studies can still be informative. Here, our classifier dealt with 14 groups and, after filtering, reached 56.3% general accuracy as well as 84% precision on biological groups. This falls within the higher range in terms of precision and number of predicted groups compared to previous studies on *in situ* images of zooplankton, especially considering that 67%–83% accuracy is often used as a benchmark for plankton classifications (Culverhouse et al., 2003; Hu and Davis, 2005). While there is still room for improvement in the original classification rates, the data filtering method presented in this study markedly improved the performance of the standard *ZooProcess/PkID* classification.

Large image datasets are likely to become increasingly common thanks to the development of affordable high-frequency, high-resolution cameras like the one installed on ISIIS. In such big datasets, all the information may not be essential and some may be efficiently omitted (Bi et al., 2015). The filtering approach used in this study considerably subsampled the data (72% of objects were discarded) in order to focus only on well-predicted objects. Despite this high subsampling rate, the two dimensional, and to a lesser extent vertical, distributions of many classes were not significantly different between the subsampled and the total, reference dataset. In addition, the poorly predicted groups could be easily identified by the sparseness of their predicted distribution and/or the high proportion of discarded images (>90%). This provided an additional control for the validation of automatically predicted distributions.

More importantly, studying realistic ecological questions with the reference and predicted datasets resulted in the same conclusions.

The size distribution of objects of most classes (9 of 12) was similarly represented in both the automatically predicted and filtered dataset (Fig. 2). In the three other classes, the filtering method discarded small objects (<5 mm) more often than larger ones, possibly because small objects are more prone to be misclassified due to their lower level of detail.

The results also highlighted the foremost influence of the frontal structure, marked by a salinity gradient, on the distributions of organisms along the across-front section (Fig. 5). This is consistent with many studies from the literature (Boucher, 1984; Goffart et al., 1995; Pedrotti and Fenaux, 1992). For example, some taxa like Acantharian radiolarians, doliolids, fish larvae, and, to a lesser extent, copepods were mostly observed in the coastal or frontal zones and in the upper 50 m of the water column (Figs. 3 and 4). Both datasets allowed us to relate the abundance of various taxa to the salinity gradient, which marks the frontal region, the intensity of the fluorescence of chlorophyll *a* associated with the DCM, or the warmer temperatures found near the surface (Fig. 5). Overall, 86% of the relationships with environmental variables that were explored were not statistically different between the two datasets. Finally, diatom chains were most abundant in the deeper layers of the central zone, where copepod concentrations were the lowest (Fig. 3), suggesting a possible influence of grazing. These results suggest that species–environment relationships or interspecific interactions can be studied at the very fine scales that imaging techniques provide without requiring labour-intensive validation.

Changes in vertical distributions between day and night, even over less than 10 m, could also be detected in the predicted data for most taxa, with a power and resolution similar to that of the reference dataset (Fig. 6; Table 4). Diel vertical migrations of copepods and medusae are well described in the literature (e.g. Hays, 2003 and Sabatés et al., 2010). However, the apparent <10 m vertical movements of solitary Colodaria radiolarians or the 2 m downward displacement of doliolids during the day are not documented in prior studies, possibly because they were missed by other sampling methods with lower vertical resolutions. The ecological significance of these fine scale vertical movements is not within the scope of this study, but the fact that they could be detected

highlights the efficacy of both high frequency imaging systems and this automatic classification and filtration method in exploring microscale processes in the plankton.

Nonetheless, some taxa share striking similarities and only a trained expert may be able to differentiate between them. These size and shape resemblances can lead to high error rates in the automatic prediction of these groups (Fernandes et al., 2009). Automatic classification methods may never reach the taxonomical resolution achieved by experts observing plankton through a stereomicroscope (even if both make mistakes; Culverhouse et al., 2003). Still, combined with data filtering, automatic classification can accurately describe spatial distributions when low taxonomical resolution is acceptable, for example to study broad groups that provide an environmental or biological context for a species of interest. Eventually, manual validation is likely to still be required in order to focus on some specific taxonomic group. For example, fish larvae imaged here were very diverse and appeared similar to appendicularians and chaetognaths in terms of body size, shape and opacity. As a result, this group was badly predicted and manual methods would still be necessary to tease apart their distribution.

Using the proposed method, the processing of 1.5 million objects required only the manual classification of 5979 objects (0.41%). It could properly describe distribution patterns, but the drastic filtering process would lead to vastly underestimating the abundances of all groups. In future studies, these underestimated abundances could be scaled up by quantifying, in each class, the proportion of discarded and wrongly classified objects (e.g. with a confusion matrix). This quantification requires to manually validate a random subset of images of each category of the predicted dataset, thus requiring additional human effort. However, during validation of the 1.5 million in this project, the throughput of a trained operator was about 10,000 objects per day. Therefore, human effort on the order of a couple of weeks would probably yield enough data to correct abundances and further control the error rate for the rest of the predicted images.

The present method is based on two features shared by all machine learning methods: the use of a learning set to teach the model how to differentiate between classes and the computation of a final score, or probability, for each object to belong in each class. The probability thresholds for the filtering step are computed by cross-validating the learning set and do not require additional manual sorting. In many cases, Random Forest, working on a few dozen features deterministically measured on the object, came out as the most efficient classifier for plankton data (e.g. Bell and Hopcroft, 2008; Fernandes et al., 2009 and Gorsky et al., 2010). Yet, overall accuracy was never more than 80%. However, deep machine learning methods such as convolutional neural networks (CNNs) are emerging as promising tools for a range of image classification tasks (Krizhevsky et al., 2012; Simonyan and Zisserman, 2015). Applying the filtering method described here to classifiers that already achieve high accuracy on large datasets may eventually lead to near-perfect automatic classifications, without discarding too much information. Such a combination would allow the handling of large plankton imaging datasets that are still challenging to process rapidly and accurately (Benfield et al., 2007; Culverhouse et al., 2006), hence providing appropriate tools to explore the finescale and microscale processes occurring in the oceans.

Acknowledgements

The authors thank A. Maupetit and F. Ferrando for their help with the manual identification, the crew of the R/V Tethys 2 operating during the *VISUFRONT* cruise and CNRS/INSU for the ship time. This work was supported by a grant from the Partner University Fund to JOI and RKC. RF's doctoral fellowship was provided by the French Ministry for Education and Research (n° 247/2012).

References

Bakun, A., 2006. Fronts and eddies as key structures in the habitat of marine fish larvae: opportunity, adaptive response. Sci. Mar. 105–122.

Belkin, I.M., 2002. Front. Interdiscip. Encycl. Mar. Sci.

Belkin, I.M., Cornillon, P.C., Sherman, K., 2009. Fronts in large marine ecosystems. Prog. Oceanogr. 81, 223–236.
Bell, J.L., Hopcroft, R.R., 2008. Assessment of ZooImage as a tool for the classification of zooplankton. J. Plankton Res. 30, 1351–1367.
Benfield, M.C., Davis, C.S., Wiebe, P.H., Gallager, S.M., Gregory Loughj, R., Copley, N.J., 1996. Video plankton recorder estimates of copepod, pteropod and larvacean distributions from a stratified region of Georges Bank with comparative measurements from a MOCNESS sampler. Deep-Sea Res. II 43, 1925–1945.
Benfield, M., Grosjean, P., Culverhouse, P., Irigolen, X., Sieracki, M., Lopez-Urrutia, A., Dam, H., Hu, Q., Davis, C., Hanson, A., Pilskaln, C., Riseman, E., Schulz, H., Utgoff, P., Gorsky, G., 2007. RAPID: Research on automated plankton identification. Oceanography 20, 172–187.
Benoit-Bird, K.J., McManus, M.A., 2012. Bottom-up regulation of a pelagic community through spatial aggregations. Biol. Lett. 8, 813–816.
Bi, H., Cook, S., Yu, H., Benfield, M.C., Houde, E.D., 2013. Deployment of an imaging system to investigate fine-scale spatial distribution of early life stages of the ctenophore Mnemiopsis leidyi in Chesapeake Bay. J. Plankton Res. 35, 270–280.
Bi, H., Guo, Z., Benfield, M.C., Fan, C., Ford, M., Shahrestani, S., Sieracki, J.M., 2015. A semi-automated image analysis procedure for in situ plankton imaging systems. PLoS One 10, e0127121.
Boucher, J., 1984. Localization of zooplankton populations in the Ligurian marine front: role of ontogenic migration. Deep-Sea Res. 31, 469–484.
Boucher, J., Ibanez, F., Prieur, L., 1987. Daily and seasonal variations in the spatial distribution of zooplankton populations in relation to the physical structure in the Ligurian Sea Front. J. Mar. Res. 45, 133–173.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Cowen, R.K., Greer, A.T., Guigand, C.M., Hare, J.A., Richardson, D.E., Walsh, H.J., 2013. Evaluation of the in situ ichthyoplankton imaging system (ISIIS): comparison with the traditional (bongo net) sampler. Fish. Bull. 111, 1–12.
Cowen, R.K., Guigand, C.M., 2008. In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. Limnol. Oceanogr. Methods 6, 126–132.
Culverhouse, P.F., Williams, R., Benfield, M., Flood, P.R., Sell, A.F., Mazzocchi, M.G., Buttino, I., Sieracki, M., 2006. Automatic image analysis of plankton: Future perspectives. Mar. Ecol. Prog. Ser. 312, 297–309.
Culverhouse, P.F., Williams, R., Reguera, B., Herry, V., González-Gil, S., 2003. Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. Mar. Ecol. Prog. Ser. 247, 17–25.
Davis, C.S., Gallager, S.M., Solow, A.R., 1992. Microaggregations of oceanic plankton observed by towed video microscopy. Science 257, 230–232.
Davis, C.S., Hu, Q., Gallager, S.M., Tang, X., Ashjian, C.J., 2004. Real-time observation of taxa-specific plankton distributions: An optical sampling method. Mar. Ecol. Prog. Ser. 284, 77–96.
Dutilleul, P., Clifford, P., Richardson, S., Hemon, D., 1993. Modifying the *t* test for assessing the correlation between two spatial processes. Biometrics 49, 305.
Espinasse, B., Zhou, M., Zhu, Y., Hazen, E.L., Friedlaender, A.S., Nowacek, D.P., Chu, D., Carlotti, F., 2012. Austral fall-winter transition of mesozooplankton assemblages and krill aggregations in an embayment west of the Antarctic Peninsula. Mar. Ecol. Prog. Ser. 452, 63–80.
Fei-Fei, L., Fergus, R., Perona, P., 2007. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. Comput. Vis. Image Underst. 106, 59–70.
Fernandes, J.A., Irigoien, X., Boyra, G., Lozano, J.A., Inza, I., 2009. Optimizing the number of classes in automated zooplankton classification. J. Plankton Res. 31, 19–29.
Gasparini, S., Antajan, E., 2013. PLANKTON IDENTIFIER: a software for automatic recognition of planktonic organisms.
Goffart, A., Hecq, J.-H., Prieur, L., 1995. Controle du phytoplancton du bassin Ligure par le front liguro-provencal (secteur Corse). Oceanol. Acta 18, 329–342.
Gorsky, G., Ohman, M.D., Picheral, M., Gasparini, S., Stemmann, L., Romagnan, J.B., Cawood, A., Pesant, S., García-Comas, C., Prejger, F., 2010. Digital zooplankton image analysis using the ZooScan integrated system. J. Plankton Res. 32, 285–303.
Greer, A.T., Cowen, R.K., Guigand, C.M., Hare, J.A., 2015. Fine-scale planktonic habitat partitioning at a shelf-slope front revealed by a high-resolution imaging system. J. Mar. Syst. 142, 111–125.
Greer, A.T., Cowen, R.K., Guigand, C.M., Hare, J.A., Tang, D., 2014. The role of internal waves in larval fish interactions with potential predators and prey. Prog. Oceanogr. 127, 47–61.
Greer, A.T., Cowen, R.K., Guigand, C.M., McManus, M.A., Sevadjian, J.C., Timmerman, A.H.V., 2013. Relationships between phytoplankton thin layers and the fine-scale vertical distributions of two trophic levels of zooplankton. J. Plankton Res. 35, 939–956.
Grimes, C.B., Finucane, J.H., 1991. Spatial distribution and abundance of larval and juvenile fish, chlorophyll and macrozooplankton around the Mississippi River discharge plume, and the role of the plume in fish recruitment. Mar. Ecol. Prog. Ser. 75, 109–119.
Hays, G.C., 2003. A review of the adaptive significance and ecosystem consequences of zooplankton diel vertical migrations. Hydrobiologia 503, 163–170.
Herman, A.W., Mitchell, M.R., Young, S.W., 1984. A continuous pump sampler for profiling copepods and chlorophyll in the upper oceanic layers. Deep-Sea Res. 31, 439–450.
Hu, Q., Davis, C., 2005. Automatic plankton image recognition with co-occurrence matrices and support vector machine. Mar. Ecol. Prog. Ser. 295, 21–31.
Irisson, J.-O., Paris, C.B., Guigand, C., Planes, S., 2010. Vertical distribution and ontogenetic migration in coral reef fish larvae. Limnol. Oceanogr. 55, 909–919.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 1–9.
Legendre, L., Le Fèvre, J., 1991. From individual plankton cells to pelagic marine ecosystems and to global biogeochemical cycles. In: Demers, S. (Ed.), Particle Analysis in Oceanography SE—11. In: NATO ASI Series, Springer, Berlin, Heidelberg, pp. 261–300.
Li, Z., Member, S., Zhao, F., Liu, J., Member, S., Qiao, Y., 2014. Pairwise nonparametric discriminant analysis for binary plankton image recognition. IEEE J. Ocean. Eng. 39, 695–701.

Lough, R.G., Broughton, E.A., 2007. Development of micro-scale frequency distributions of plankton for inclusion in foraging models of larval fish, results from a Video Plankton Recorder. J. Plankton Res. 29, 7–17.
Luo, J., Grassian, B., Tang, D., Irisson, J., Greer, A., Guigand, C., McClatchie, S., Cowen, R., 2014. Environmental drivers of the fine-scale distribution of a gelatinous zooplankton community across a mesoscale front. Mar. Ecol. Prog. Ser. 510, 129–149.
MacLeod, N., Benfield, M., Culverhouse, P., 2010. Time to automate identification. Nature 467, 154–155.
McClatchie, S., Cowen, R., Nieto, K., Greer, A., Luo, J.Y., Guigand, C., Demer, D., Griffith, D., Rudnick, D., 2012. Resolution of fine biological structure including small narcomedusae across a front in the Southern California Bight. J. Geophys. Res. 117, C04020.
Neilson, J.D., Perry, R.I., 1990. Diel vertical migrations of marine fishes: an obligate or facultative process? Adv. Mar. Biol. 26, 115–168.
Olson, D.B., Hitchcock, G.L., Mariano, A.J., Ashjian, C.J., Peng, G., Nero, R.W., Podesta, G.P., 1994. Life on the edge: marine life and fronts. Oceanography 7, 52–60.
Pedrotti, M.L., Fenaux, L., 1992. Dispersal of echinoderm larvae in a geographical area marked by upwelling (Ligurian Sea, NW Mediterranean). Mar. Ecol. Prog. Ser. 86, 217–227.
Picheral, M., Guidi, L., Stemmann, L., Karl, D., Iddaoud, G., Gorsky, G., 2010. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. Limnol. Oceanogr. Methods 8, 462–473.
Pinel-Alloul, B., 1995. Spatial heterogeneity as a multiscale characteristic of zooplankton community. In: Space Partition within Aquatic Ecosystems. Springer, pp. 17–42.
Sabatés, A., Pagès, F., Atienza, D., Fuentes, V., Purcell, J.E., Gili, J.M., 2010. Planktonic cnidarian distribution and feeding of Pelagia noctiluca in the NW Mediterranean Sea. Hydrobiologia 645, 153–165.
Sammari, C., Millot, C., Prieur, L., 1995. Aspects of the seasonal and mesoscale variabilities of the Northern Current in the western Mediterranean Sea inferred from the PROLIG-2 and PROS-6 experiments. Deep-Sea Res. I 42, 893–917.
Samson, S., Hopkins, T., Remsen, A., Langebrake, L., Sutton, T., Patten, J., 2001. A system for high-resolution zooplankton imaging. IEEE J. Ocean. Eng. 26, 671–676.
Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recoginition. Int. Conf. Learn. Represent. 1–14.
Smeti, H., Pagano, M., Menkes, C., Lebourges-Dhaussy, A., Hunt, B.P., Allain, V., Rodier, M., de Boissieu, F., Kestenare, E., Sammari, C., 2015. Spatial and temporal variability of zooplankton off New Caledonia (Southwestern Pacific) from acoustics and net measurements. J. Geophys. Res. Oceans 120, 1–25.
Solow, A.R., Bollens, S.M., Beet, A., 2000. Comparing two vertical plankton distributions. Limnol. Oceanogr. 45, 506–509.
Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging in-flow cytometry. Limnol. Oceanogr. Methods 5, 204–216.
Vandromme, P., Stemmann, L., Berline, L., Gasparini, S., Mousseau, L., Prejger, F., Passafiume, O., Guarini, J.-M., Gorsky, G., 2011. Inter-annual fluctuations of zooplankton communities in the Bay of Villefranche-sur-mer from 1995 to 2005 (Northern Ligurian Sea, France). Biogeosciences 8, 3143–3158.
Wiebe, P.H., Benfield, M.C., 2003. From the Hensen net toward four-dimensional biological oceanography. Prog. Oceanogr. 56, 7–136.
Ye, L., Chang, C.Y., Hsieh, C.H., 2011. Bayesian model for semi-automated zooplankton classification with predictive confidence and rapid category aggregation. Mar. Ecol. Prog. Ser. 441, 185–196.
Zhao, F., Lin, F., Seah, H.S., 2010. Binary SIPPER plankton image classification using random subspace. Neurocomputing 73, 1853–1860.