Full length article

The LOKI underwater imaging system and an automatic identification model for the detection of zooplankton taxa in the Arctic Ocean
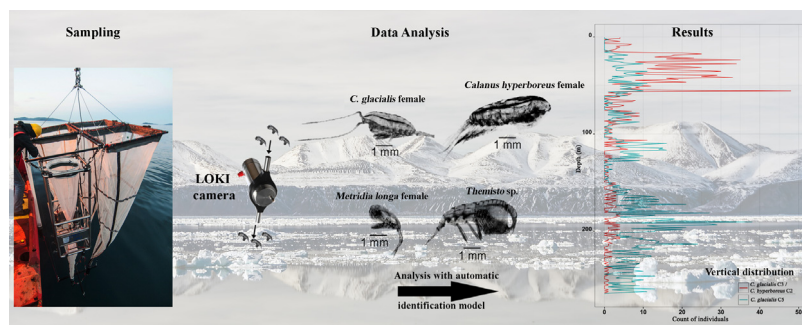
Moritz Sebastian Schmid *, Cyril Aubry, Jordan Grigor, Louis Fortier

Takuvik Joint International Laboratory, Université Laval, Québec, QC, Canada G1V 0A6
Québec-Océan, Université Laval, Québec, QC, Canada G1V 0A6
Département de Biologie, Université Laval, Québec, QC, Canada G1V 0A6

GRAPHICAL ABSTRACT

HIGHLIGHTS

- An automatic zooplankton identification model has been developed for 114 taxonomic categories.
- The model successfully distinguishes species and stages.
- Various model validations show high model performance for identifying key zooplankton taxa.
- The model makes unprecedented insights into the fine scale vertical distribution of taxa possible.

* Corresponding author at: Takuvik Joint International Laboratory, Université Laval, Québec, QC, Canada G1V 0A6.
E-mail addresses: Moritz.Schmid@takuvik.ulaval.ca (M.S. Schmid), Cyril.Aubry@qo.ulaval.ca (C. Aubry), JordanGrigor@gmail.com (J. Grigor), Louis.Fortier@bio.ulaval.ca (L. Fortier).

A B S T R A C T

We deployed the Lightframe On-sight Keyspecies Investigation (LOKI) system, a novel underwater imaging system providing cutting-edge imaging quality, in the Canadian Arctic during fall 2013. A Random Forests machine learning model was built to automatically identify zooplankton in LOKI images. The model successfully distinguished between 114 different categories of zooplankton and particles. The high resolution taxonomical tree included many species, stages, as well as sub-groups based on animal orientation or condition in images. Results from a machine learning regression model of prosome length ($R^2 = 0.97$) were used as a key predictor in the automatic identification model. Model internal validation of the automatic identification model on test data demonstrated that the model performed with overall high accuracy (86%) and specificity (86%). This was confirmed by confusion matrices for external testing results, based on automatic identifications for 2 complete stations. For station 101, from which images had also been used for training, accuracy and specificity were 85%. For station 126, from which images had not been used to train the model, accuracy and specificity were 81%. Further comparisons between model results and microscope identifications of zooplankton in samples from the two test stations were in good agreement for most taxa. LOKI's image quality makes it possible to build accurate automatic identification models of very high taxonomic detail, which will play a critical role in future studies of zooplankton dynamics and zooplankton coupling with other trophic levels.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Zooplankton are a diverse group of animals with limited swimming ability inhabiting aquatic ecosystems worldwide. They transfer energy and carbon from autotrophic organisms to organisms higher up the trophic chain, such as forage fish and seabirds (e.g. Bradstreet and Cross, 1982 and Welch et al., 1992). The Arctic is meanwhile affected by rapid climate change (Overland et al., 2014; Serreze and Francis, 2006), including important changes in the phenology of primary producers (Ardyna et al., 2014; Arrigo et al., 2008). Here, changes in the distributions, abundances and behaviour of zooplankton are being observed (Lischka and Riebesell, 2012; Richardson, 2008; Wassmann et al., 2011) potentially leading to major ecosystem shifts (Grebmeier et al., 2006; Falk-Petersen et al., 2007; Søreide et al., 2010). Effective monitoring of zooplankton population dynamics and their interactions with primary producers is therefore critical.

Although zooplankton sampling still relies heavily on nets due to their inexpensiveness and ease of deployment, new *in-situ* imaging systems for zooplankton have been developed since the 1980s (Davis et al., 1992b; Jaffe et al., 2001; Ortner et al., 1979; Wiebe and Benfield, 2003). The main advantage of *in-situ* imaging systems over nets is the unprecedented high spatial resolution of the data obtained (Davis et al., 1992a), permitting the fine scale study of zooplankton distribution along hydrographic gradients such as fronts and clines (e.g., Haury et al., 1978 and Valiela, 1995). Each imaging system has different qualities and characteristics based on the chosen optical setup and implementation (see Schulz et al., 2010 and Schulz, 2013). For instance, the volume of water imaged varies between samplers, and is important in determining the representativeness of images for abundance estimation (Davis et al., 1992b; Benfield et al., 1996; Cowen and Guigand, 2008). Amongst the few available *in-situ* imaging systems, the recently developed Lightframe On-sight Keyspecies Investigation (LOKI) system is unique

in that it includes a net and cod end to concentrate and capture zooplankton in a constrained volume of water, so that imaged specimens can be analysed further in the laboratory (Schulz et al., 2010; Hirche et al., 2014). Since underwater imaging systems often return thousands of images from a single haul, it has been desirable to develop and apply methods of automatically identifying taxa in images (Rolke and Lenz, 1984). Zooplankton identification from images has become a public endeavour in recent years (Zooniverse, www.planktonportal.org), and development of automatic identification models has been subject to prominent programming competitions (National Science Bowl, www.kaggle.com). In contrast to manual identification, automatic identification models can analyse every zooplankton individually instead of relying on extreme subsampling (Guelpen et al., 1982). Also, the model can quickly be applied to new data to increase analytical capacity (Benfield et al., 2007) and the error for each category in the model is known, constant, and may be corrected for (Solow et al., 2001). Therefore, automatic approaches address poor consistency and accuracy of manual identifications (Culverhouse et al., 2003, 2014). Modern automatic identification models are based on machine learning (ML) algorithms, whose underlying concept is that a machine can "learn", in many small steps, without being explicitly programmed for each decision taken (Kovahi and Provost, 1998). ML algorithms are effective for detecting patterns in data and making predictions based on them (Strobl et al., 2009). Algorithms currently in use for automatic zooplankton identification include support vector machines (Cortes and Vapnik, 1995), Convolutional Neural Networks (Fukushima, 1980), and the decision tree based ensemble model Random Forests (RF, Breiman, 2001). However, models successfully implemented in this field have lacked taxonomic detail (usually with only 10–20 categories), and quality of zooplankton images taken by optical imagers has only recently reached a level where body structures of zooplankton can be seen (often key to identifying them). This study used high resolution *in-situ* images of zooplankton in the Canadian Arctic taken by LOKI to develop a highly specific automatic zooplankton identification model. The ability of the newly developed model to identify taxa to a very high taxonomic level was tested by feeding the model new images.

2. Materials and methods

2.1. The LOKI system

The Lightframe On-sight Keyspecies Investigation (LOKI) system consists of 4 main units: (1) A plankton concentration net, with mouth opening of 0.28 m^2 and a mesh size of 200 μm, (2) the LOKI computer with various environmental sensors, (3) the camera system, and (4) the battery (Fig. 1(a)). The system configuration is designed for vertical towing in the water column. Zooplankton and particles enter the concentration net from the top, before flowing through a channel that passes in front of the camera (Fig. 1(b), (c)), a Prosilica GC 1380H camera (AVT—Allied Vision Technologies, Canada) with a Pentax 2514-M lens and an image resolution of 1360 × 1024 pixel at 30 fps. Dark field imaging and an image resolution of 23 μm per pixel were used here. A high power LED flash unit, synchronized with the exposure trigger signal of the camera, allows for a fast shutter time (55 μs), avoiding motion blurring and image distortion (see Schulz, 2013 for more details). In combination with an only 4 mm high imaging channel (length = 31.3 mm, width = 20.75 mm, volume = 2.6 cm^3) this leads to all imaged particles being in focus. Single organisms are detected and cut out from the field of view of the camera and stored on the solid state drive (SSD) of LOKI, where they are later accessed for further analyses.

2.2. LOKI deployment

Sampling was conducted during the ArcticNet 2013 expedition (July 27 to September 8) onboard the CCGS *Amundsen*. Vertical tows of LOKI were conducted in the Northwest Passage (74°N 95°W), northern Baffin Bay and the North Water Polynya (NOW, 77°N 75°W) and northward to the vicinity of the Peterman Glacier (81°N 62°W). Imaged zooplankters were collected in a custom-made cod end, and upon retrieval were preserved in 4% formaldehyde–seawater solution buffered with sodium-borate. Sensors attached to LOKI recorded temperature and oxygen (Aanderaa Oxygen Optode 4330F),
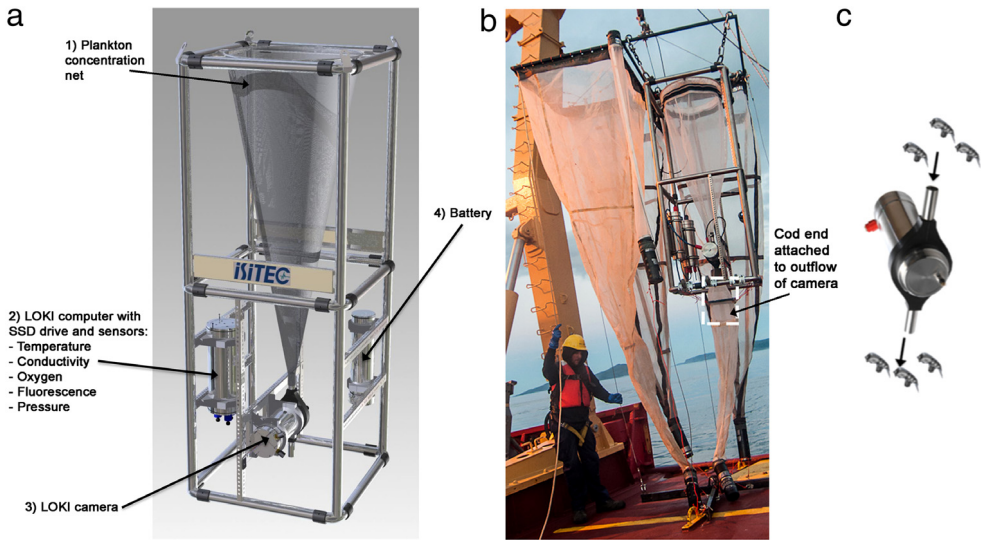
Fig. 1. (a) Schematic of LOKI showing its main components 1–4. (b) The LOKI system on the right attached to a frame besides a traditional zooplankton net sampler, during a recent deployment in the Canadian Arctic. (c) The LOKI camera, showing how plankton passes through the channel for imaging. (a) and (c) are adapted from Isitec GmbH. Photo credit for (b): Jessy Barrette.

conductivity (Aanderaa Conductivity Sensor 3919), fluorescence (TriOS MicroFlu-chl) and pressure (Aanderaa Pressure Sensor 4017D) per second during each haul.

2.3. Preparing LOKI data for building prosome length, prosome width and automatic identification models

A total of 15 hauls from 9 stations across the Canadian Arctic were used to source images for this study (Fig. 2, Table 1). Prosome length, width, and automatic identification models were trained on images from 14 hauls, ensuring that all zooplankton taxa likely to be encountered in the study area were included in the three models. Images from a single haul were used exclusively in the testing of the automatic identification model (Fig. 2, Table 1).

For each haul, all LOKI images were imported into a specific software designed to analyse LOKI images and build a zooplankton classification tree (LOKI browser; Schulz et al., 2010). The LOKI browser measuring function, with a threshold of 15 on a scale from 0 (black) to 255 (white), was used to measure image parameters (hereafter referred to as image feature set 1, Table 2), such as object area (mm^2), circularity (Wojnar and Kurzydlowski, 2000), Hu moments (Hu, 1962) and Fourier descriptors (Zhang and Lu, 2002) from the 8 bit greyscale images. The ZOOMIE v 1.0 (Zooplankton Multiple Image Exclusion) software (Schmid et al., 2015) was applied to all images. Written in PhP and JavaScript, ZOOMIE ensures that LOKI data can be considered representative of the plankton present by detecting multiple images of a single individual (hereafter referred to as "double images"), which should be removed from the analyses. Double images are taken when flow speed inside the camera channel is reduced, primarily due to changes in the pressure at the entrance of the imaging channel due to e.g. aggregations of biomass or additional heaving of the sampling platform. Each image was assigned the *x*- and *y* pixel position which it had when cut out of the overall camera frame during deployment. An image was flagged as a double image when its *x* and *y* positions changed from the previous image only by a limited number of pixels (e.g. 80 pixels) during a short time period, and when a set of ten rules was unanimously fulfilled. Each rule was based on a different image parameter (e.g. area in mm^2, greymean and kurtosis, see Schmid et al., 2015 for details). Image artefacts that could prevent correct zooplankton identification such as smears on the camera lens or light reflections on image borders, were detected and deleted in Adobe Photoshop CS6. "Clean" images were re-imported
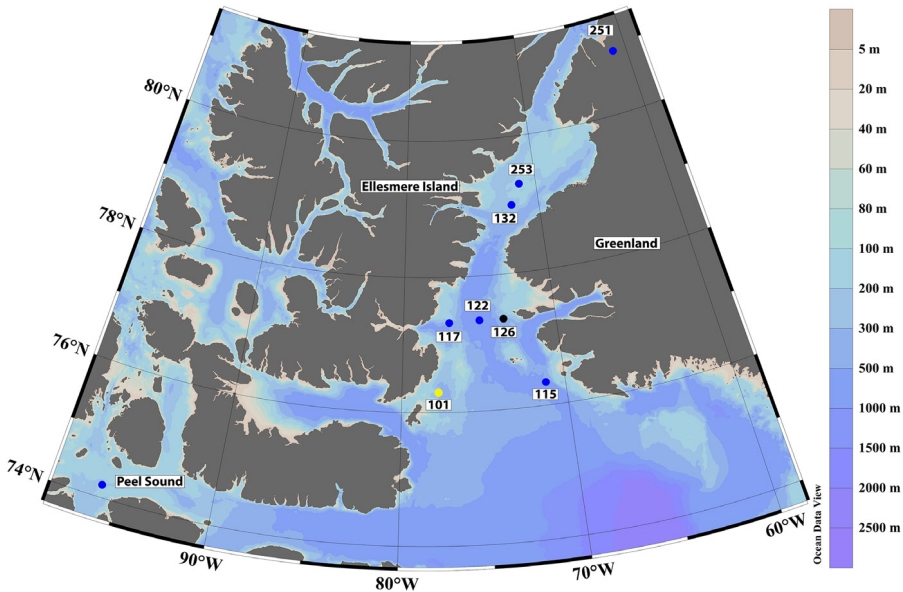
Fig. 2. Stations in the Canadian Arctic used to train (blue points) and test (black points) the models developed here. Images from station 101 (yellow points) were used in both training and testing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Stations included in training and testing the models developed in this study.

Station	Date	Latitude °N	Longitude °W	Bottom depth (m)	Usage	Total number of images collected	Number of images used in training
101	2013-08-15	76°23.23	77°22.98′	349	Training, testing	9 418	1412
"	2013-08-15	76°22.67′	77°22.93′	357	Training	18 772	1161
"	2013-08-15	76°20.71′	77°34.47′	356	Training	10 781	304
"	2013-08-16	76°19.78′	77°42.2′	246	Training	10 644	291
"	2013-08-16	76°17.48′	77°45.62′	282	Training	13 499	303
115	2013-08-18	76°20.477′	71°11.756′	654	Training	15 632	423
"	2013-08-18	76°21.6′	71°12.54′	631	Training	24 347	534
132	2013-08-20	78°59.94′	72°3.7′	216	Training	1 956	252
251	2013-08-22	81°12.40′	62°07.28′	835	Training	6 421	281
253a	2013-08-25	79°17.73′	71°17.67′	190	Training	4 113	249
"	2013-08-25	79°17.513′	71°17.911′	181	Training	2 063	250
122	2013-08-27	77°20.691′	75°0.976′	641	Training	21 521	507
126	2013-08-27	77°20.65′	73°25.64′	324	Testing	4 083	na
117	2013-08-28	77°19.45′	77°0.47′	449	Training	2 625	300
Peel sound	2013-09-02	74°11.23′	95°56.29′	192	Training	4 904	568
"	2013-09-02	74°10.327′	95°48.307′	189	Training	5 181	418

and image feature set 1 was re-measured in LOKI browser to account for changing parameters after artefact removal.

Categories of zooplankton in the classification tree were chosen as taxonomically detailed as possible and included species and copepod stages (Fig. 3), but also orientational subgroups, as the same individual could appear very differently in images depending on its position (see Fig. 4). Zooplankton from each taxonomic/orientational category were manually identified from images with LOKI browser. Images were selected from depths across the whole water column to avoid selection

Table 2
Image features used in training the machine learning models for prosome length, prosome width and automatic species identification. Usage indicates the models: L indicates the prosome length model, W the prosome width model and AI the automatic species identification model. Image feature set 1 originates from LOKI browser and image feature set 2 originates from GUIDOS toolbox.

No	Image feature	Image basis	Explanation	Usage	Image feature set	Reference
1	Area in mm²	Greyscale	Area of the object in mm²	L/W, AI	1	–
2	Convexity	Greyscale	Degree of outward curvature	L/W, AI	1	Berger (1990)
3	Form	Greyscale	Values of ≥ 1. A value of 1 resembles a circle	L/W, AI	1	Wojnar and Kurzydlowski (2000)
4–13	Fourier descriptors 1–10	Greyscale	Contour based shape descriptors based on Fourier transformations	L/W, AI	1	Zhang and Lu (2002)
14	Greymean	Greyscale	Mean grey value of pixels	L/W, AI	1	–
15	Homogeneity	Greyscale	Based on the coocurrence matrix of grey values	L/W, AI	1	Knauer and Meffert (2010)
16–22	Hu moments 1–7	Greyscale	Invariant image moments (moment = particular weighted average of the image pixels' intensities)	L/W, AI	1	Hu (1962)
23	Kurtosis	Greyscale	Tail extremity of grey values	L/W, AI	1	Westfall (2014)
24	Skewness	Greyscale	Grey value distribution	L/W, AI	1	Von Hippel (2005)
25–27	MSPA Core	Binary	Core pixels are defined as foreground pixels (value 2 in the binary image) with a distance to background pixels (value 1 in the binary image) that is greater than the specified edge width value (8 pixels in this case)	L/W, AI	2	Soille and Vogt (2009)
28–30	MSPA Islet	Binary	Islet pixels are foreground connected components which are too small to contain core pixels	L/W, AI	2	Soille and Vogt (2009)
31–33	MSPA Loop	Binary	Connectors: Foreground pixels linking back to or traversing a perforation within the same core component	L/W, AI	2	Soille and Vogt (2009)
34–36	MSPA Bridge	Binary	Connectors: Foreground pixels linking two or more different core components	L/W, AI	2	Soille and Vogt (2009)
37–39	MSPA Perforation	Binary	Boundaries: Foreground pixels surrounding a hole and forming an internal object boundary with a distance equal to the specified edge width value	L/W, AI	2	Soille and Vogt (2009)
40–42	MSPA Edge	Binary	Boundaries: Foreground pixels surrounding a core and forming an external object boundary with a distance equal to the specified edge width value	L/W, AI	2	Soille and Vogt (2009)
43–45	MSPA Branch	Binary	Do not belong to any of the above categories. They originate either from edges, perforations, bridges or loops	L/W, AI	2	Soille and Vogt (2009)
46–47	MSPA Background	Binary	All background pixels	L/W, AI	2	Soille and Vogt (2009)

Table 2 (*continued*)

No	Image feature	Image basis	Explanation	Usage	Image feature set	Reference
48	Meshsize	Binary	Probability of two points (pixels) chosen randomly within an area being connected. Equivalent to the sum of squares divided by the image area	*L/W*, AI	2	Jaeger (2000)
49	Computation time meshsize	Binary	Time (seconds) it takes to compute the meshsize value	*L/W*, AI	2	Jaeger (2000)
50	Relative fragmentation	Binary	Meshsize of the actual image scaled into the range of min and max meshsize	*L/W*, AI	2	Jaeger (2000)
51	Predicted prosome length	Derived	Values of predicted prosome length	AI	–	This publication
52	Predicted prosome width	Derived	Values of predicted prosome length	AI	–	This publication

Features belonging to MSPA analysis represent 3 features: (1) Foreground (%), (2) All data (%) and (3) Frequency. E.g. MSPA Core is divided into (1) Core foreground (%), the portion of foreground pixels classified as core, (2) Core data (%), the portion of all pixels classified as core, and (3) Frequency: the number of core components classified in the image. Only MSPA Background consists of 2 features, all data (%) and frequency.

bias. Of 14 558 manually identified images from the 14 hauls, 7252 were randomly selected using the sample function in R to train the model (the remaining 7306 reserved for model internal testing). Manual image identifications were performed in the Fortier laboratory at U. Laval by authors CA and MS. Due to morphological similarities and size overlaps, the young stages of the abundant copepods *C. glacialis* C2/*C. hyperboreus* C1, as well as *C. glacialis* C3/*C. hyperboreus* C2, were treated as single categories in the training database.

Lengths and widths (μm) of 5165 zooplankton specimens across all categories of the classification tree were measured with a digital ruler in LOKI browser. For copepods, prosome length and prosome width were measured. For organisms with no prosome (e.g. jellyfish) maximum length and width were taken.

For all manually classified images selected for training, image feature set 1 was re-analysed using LOKI-browser with different grey level thresholds (levels: 8, 15, 20, 35, 60), to account for variations in the appearances of a zooplankter in images. Thresholds ranged from 8, where a greater part of the image area was included in the analysis, to 60, which resulted in a more restricted thumbprint (Fig. 5). This process generated a total augmented training dataset of 36 260 images, ranging from $n = 30$ for *C. hyperboreus* males of lateral view, to $n = 1810$ for copepod nauplii.

Images were again thresholded, this time in Adobe Photoshop CS6, with grey levels: 8, 15, 20, 35, 60 and the resulting images were converted to binary images. Two further analyses were carried out on these binary images; Morphological Spatial Pattern Analysis (MSPA) and mesh size analysis, both part of GUIDOS toolbox (Graphical User Interface for the Description of image Objects and their Shapes (GUIDOS; Vogt, 2014)). GUIDOS toolbox was programmed in the IDL programming language and is a self-contained suite of raster image processing routines. MSPA describes geometry and connecting pathways in images and is solely based on geometric concepts. During the MSPA analysis process, the foreground area of an image is divided into 7 MSPA classes (Core, Islet, Perforation, Edge, Loop, Bridge and Branch; see Table 2, Fig. 6 as well as Soille and Vogt, 2009 for more details). Mesh size analysis is a fragmentation measure introduced by Jaeger (2000), which is well suited for comparing fragmentation of images with different total size. MSPA and meshsize analyses led to an additional 23 and 3 image features respectively, hereafter referred to as image feature set 2 (Table 2).

Image feature set 2 was merged with data from LOKI browser, consisting of the environmental information (measured when each image was captured), the manual classification and image feature set 1. The resulting table provided the basis for building the prosome length and width models.

2.4. Training models with Random Forests

Salford Systems, Inc. predictive modeler (version 7.0) implementation of the Random Forests (RF) algorithm was used to build all three models produced in this study. RF belongs to the family of ensemble decision tree methods and employs bootstrap aggregating (bagging; Breiman, 2001), i.e. selecting a random subset of data for each decision tree built in the ensemble of trees (the forest; Breiman, 1996). RF differs from other algorithms in that it also selects a random subset of predictors for splitting at each node in the decision tree. Trees in a RF are grown to their full size, without employing pruning. These features make RF relatively resistant against overfitting (Breiman, 2001).

RF was used to build regression models of prosome length and prosome width using image feature sets 1 and 2 as predictor variables (Table 2). Both regression models were internally tested by RF using 20% withheld test data and built on 500 trees. The best performing settings differed only in the number of predictors considered at each node (Table 3). Linear regressions for predicted prosome length and width versus their measured counterpart were carried out to investigate their accuracy. The prosome length and width models were then applied to the complete dataset of manually identified zooplankton images and a predicted prosome length and width obtained for all of them. These two measures were then included as predictors in the automatic zooplankton identification model.

Backward stepwise model selection was used for each model to sequentially remove lowest ranking predictors. If the model improved, the predictor was dropped and the model was rerun on the remaining predictors.
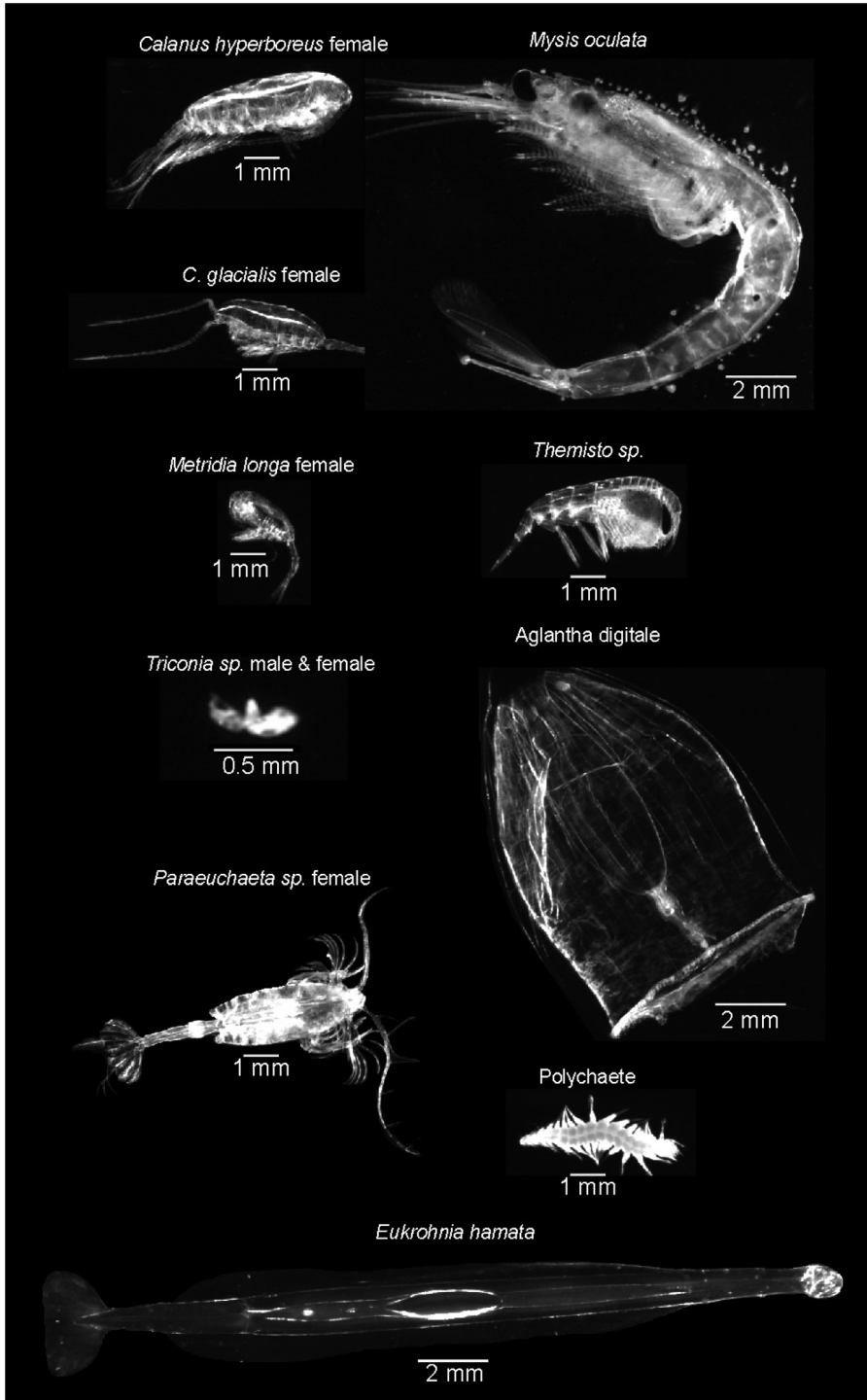
Fig. 3. LOKI images of selected zooplankton taxa.
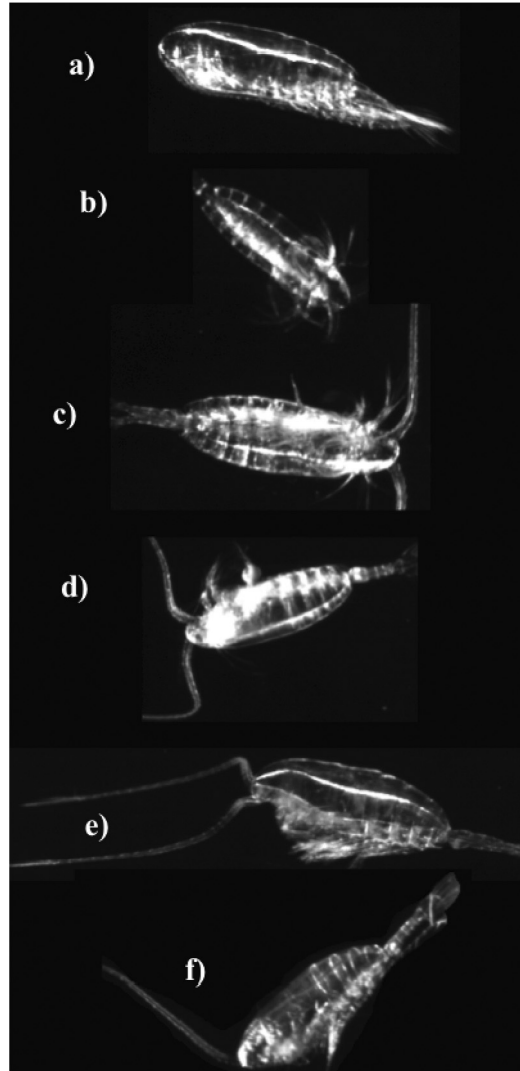
Fig. 4. Orientational subgroups were also included in the automatic identification model. (a)–(f) show females of *Calanus glacialis* in different positions (a) lateral position, (b) dorsal short position, (c)–(d) dorsal long position, (e)–(f) antenna in front position. To recognize that these images show the same species and stage required the model to be trained on orientational categories.

Table 3
Important Random Forests model settings for the regression models of prosome length and width as well as the zooplankton classification model.

Model	Type	Validation	Trees grown	Predictors at each node	Class weights
Prosome length	Regression	20% withheld test data	500	20	–
Prosome width	Regression	20% withheld test data	500	15	–
Zooplankton	Classification	Dataset of 7306 images	500	30	Balanced

Fig. 5. The copepod *Heterorhabdus* sp. measured at different thresholds. The green outline delineates the area included by LOKI browser for image feature extraction. An identification model based on these five thresholds is more robust than when just one threshold is used. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
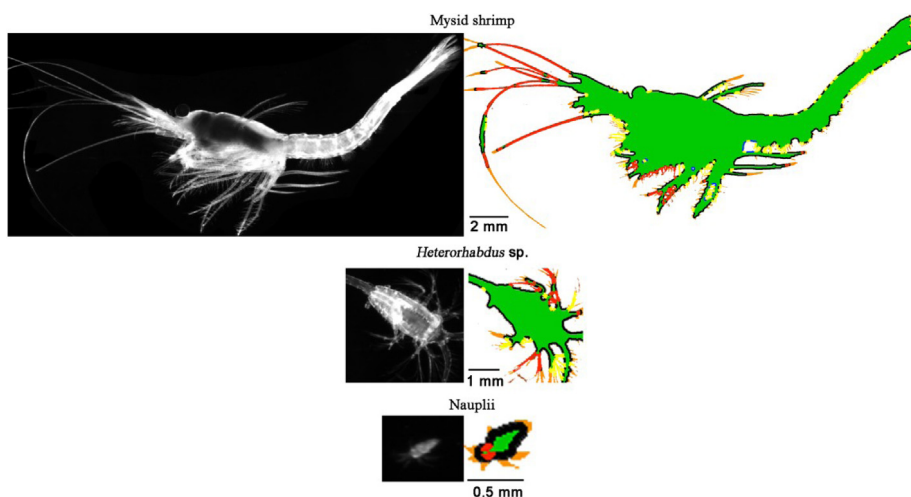

Fig. 6. Raw LOKI images on the left compared to their corresponding MSPA images on the right. The MSPA classes core (green), edge (black), bridge (red), branch (orange) and loop (yellow) each characterize different structures (e.g., maxillipeds, antennae) of the zooplankton individuals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The automatic identification model used balanced class weights causing Random Forests to account for uneven sample sizes by upweighing rare categories (i.e., incurring a higher error for misclassifying rare cases, Table 3). The classification model was also built on 500 trees but the best results were obtained using a subset of 30 predictors at each node.

2.4.1. Model internal testing

The manually identified images not used in training were used for model internal testing the automatic identification model (Table 3). The test dataset had the same underlying species distribution and images were processed in the same way as training images with the sole exception that they were only analysed based on grey level threshold 15 rather than the five used for the training images. Test sample size varied considerably between zooplankton categories (Tables A.1 and A.2), but always represented ∼20% of the augmented training sample size (ranging from $n = 6$ for rare *C. hyperboreus* males to up to $n = 429$ for *Triconia* sp.). RF calculated the specificity (%, true positive samples/classifier prediction total) for each category in the classification tree. Accuracy (%, true positive samples/total manual classification) and specificity without orientational misclassifications were calculated (e.g., *C. glacialis* C3 in lateral orientation vs. *C. glacialis* C3 in dorsal orientation). In a second step categories differing only in orientation but showing the same taxa were pooled.

2.5. External validations based on test stations

At stations 101, western NOW, and 126, eastern NOW (Fig. 2, Table 1), the developed automatic identification model was used to identify all images to further validate the model.

Automatic identifications of single images were compared with manual identifications. This was done randomly within each category of the classification tree until 50% of all images were validated and for a minimum of $n = 15$. Accuracy and specificity for each category were calculated.

Apart from a reliably working automatic identification another potential source that can bias abundance estimates is the *in-situ* imaging process itself. To test whether the LOKI system imaged the zooplankton community representatively rather than testing the model itself, the cod-end samples taken were also analysed under a binocular microscope by an expert taxonomist. Three zooplankton subsamples were taken from the cod-end of each station and 1000 copepods from each subsample were identified. Numbers of all zooplankton were recorded. Results were compared with automatic identifications.

3. Results

A total of 114 zooplankton/particle categories were included in the automatic identification model. That number was reduced to 63, for display, after assigning categories describing animal conditions or orientations to taxa. These 63 compressed categories were separated into 5 groups: (a) copepods identifiable to the stage level (typically *Calanus* copepods), (b) copepods not identifiable to the stage level, (c) gelatinous zooplankton, (d) non-gelatinous-non-copepod zooplankton and (e) other particles such as zooplankton faecal pellets (Table A.2).

3.1. Model internal testing

Based on the 114 original categories, the classifier showed average accuracy (true positive samples/total manual classification) and specificity (true positive samples/classifier prediction total) values of 85% (Table A.1), whilst these measures for the 63 compressed categories were 86% and 86% respectively.

During model internal testing, accuracy ranged from 62% for *C. glacialis* C2/*C. hyperboreus* C1 to 100% for several taxa including *C. hyperboreus* males (Table A.2). Specificity ranged from 40% for *Paraeuchaeta* sp. C3 to 100% for *Heterorhabdus* sp. The confusion matrix for model internal testing similarly revealed the high performance of the model and a low probability of misclassifications (Fig. 7). Mean probability of misclassification across all categories was 2.9%. The top 6 most important predictors in the automatic identification model were Hu moment 1, 4 and 2, followed by predicted prosome length, form and predicted prosome width (Table A.3).

The regression models for prosome length and prosome width performed very well (R^2 values of 0.97 and 0.96 for prosome length and width respectively, Fig. 8). For both models the area of the imaged organism was the most important predictor (Table A.3). Backward stepwise model selection, which was used for all models, did not lead to increased performance and therefore all models were built using all predictors.

3.2. External validations based on test stations

Automatic identifications showed that the most abundant taxa at the two stations was *C. hyperboreus* C4 (2697 identified images at station 101 and 1109 images at station 126) followed by *C. glacialis* C5 (1247 images at station 101 and 428 at station 126) and *M. longa* females (624 images at station 101 and 384 images at station 126, Table A.4).

Manual verifications of images confirmed the high model specificity and accuracy shown by internal testing. Inspections of 4820 images from station 101 and comparison with predictions revealed weighted mean accuracy and specificity of 85%. For station 126 (2163 images) weighted accuracy and specificity were at 81% each (Table 4).

Fig. 7. Results of internally testing the automatic identification model. The confusion matrix shows probabilities of classifications for the 63 compressed categories (n = 7306). Probabilities range from classifications rarely occurring (<0.01, blue) to happening regularly (1, red). White cells denote cases where no classifications occurred. High values on the diagonal indicate that the detection probability for a species was high. Low values off the diagonal delineate category-specific misclassifications. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Accuracy (Ac) and specificity (Sp) for each of the 63 compressed categories at the two test stations.

Taxon or particle	Stage	Station 101			Station 126		
		Sample size (n)	Ac (%)	Sp (%)	Sample size (n)	Ac (%)	Sp (%)
(a) Copepods identified to the stage level							
C. glacialis	C1	49	78	75	11	73	73
	C4	159	82	78	94	77	74
	C5	624	85	85	211	78	76
	F	40	81	72	23	76	71
C. hyperboreus	C3	110	82	72	40	80	72
	C4	1349	87	92	555	85	89
	C5	296	88	83	97	81	71
	F	61	87	77	68	83	84
	M	5	80	67	1	100	50
C. glacialis C2/C. hyperboreus C1	–	15	73	80	7	71	71
C. glacialis C3/C. hyperboreus C2	–	145	83	81	64	79	79
M. longa	C1	128	78	79	26	71	74
	C2	50	83	73	15	73	76
	C3	67	84	93	16	81	71
	C4	164	85	89	77	80	83
	C5	42	82	76	15	75	75
	F	312	87	89	192	82	86
	M	22	91	71	63	84	80
Paraeuchaeta sp.	C2	1	100	25	0	–	0
	C3	7	86	60	0	–	0
	C4	15	88	70	12	83	71
	C5	14	79	65	11	73	67
	F	2	100	50	2	100	33
	F with egg	0	–	0	13	77	77
Pseudocalanus sp.	C1	3	100	30	0	–	0
	C4	1	100	33	4	75	60
	C5	71	85	87	72	80	84
	F	15	83	70	5	80	67
(b) Copepods not identified to the stage level							
Aetideidae	–	8	88	54	15	87	76
Copepoda egg	–	171	99	99	16	94	94
Copepoda nauplii	–	116	84	88	58	80	81
Harpacticoida	–	7	86	75	5	80	67
Heterorhabdus sp.	–	5	80	67	2	100	100
Microcalanus sp.	–	15	78	70	1	100	20
Oithona sp.	–	20	79	78	26	76	80
Scaphocalanus sp.	–	1	100	50	8	75	75
Triconia sp.	–	291	81	88	146	77	85
(c) Gelatinous zooplankton							
Aglantha digitale	–	28	89	82	31	85	82
C. limacina adult	–	9	78	78	0	–	–
C. limacina larvae	–	29	74	77	15	74	77
Chaetognatha	–	15	78	75	6	83	63
Ctenophora	–	15	87	76	11	82	69
Limacina helicina	–	40	82	84	15	81	76
Oikopleura sp.	–	15	72	68	5	60	50
Other hydromedusae	–	5	100	83	2	100	67
Siphonophora	–	5	80	50	6	67	80

Table 4 (*continued*)

| Taxon or particle	Stage	Station 101			Station 126		
		Sample size (*n*)	Ac (%)	Sp (%)	Sample size (*n*)	Ac (%)	Sp (%)
Veliger larvae	–	15	80	70	16	74	72
(d) Non-gelatinous-non-copepod zooplankton							
Aphroditiformia larvae	–	11	73	67	10	70	70
Cirripedia cypris	–	31	94	91	2	50	33
Cirripedia nauplii	–	10	80	67	1	100	33
Echinodermata larvae (bell-shaped)	–	15	81	71	4	75	60
Echinodermata larvae (triangular-shaped)	–	15	86	77	2	50	33
Gammaridae	–	1	100	33	2	50	50
Mysidae	–	3	100	75	4	100	80
Ostracoda	–	20	85	75	13	77	71
Other amphipoda	–	15	85	71	2	100	67
Polychaeta adult	–	16	97	78	8	88	78
Polychaeta larvae	–	27	85	73	9	78	64
Radiolaria	–	27	85	84	15	80	80
Themisto sp.	–	15	95	70	4	75	75
(e) Other particles							
Detritus	–	15	79	72	2	100	40
Faecal pellet	–	15	79	76	0	–	0
Fibre	–	15	75	63	15	74	74
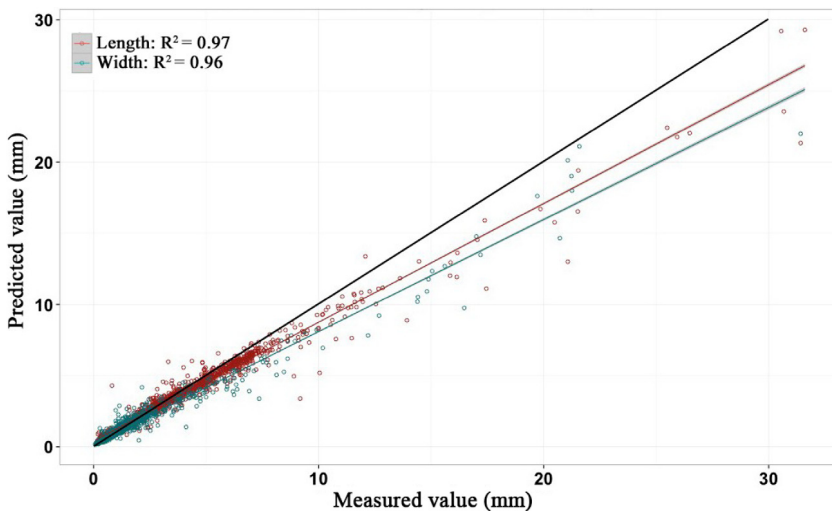
Fig. 8. Prosome lengths and prosome widths predicted by the model, and plotted against their measured values. Red and green lines are linear regressions of prosome length and width respectively. The black line indicates perfect agreement. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

For specific taxa, accuracy at station 101 ranged from 62% for *Oikopleura* sp. and 100% for several taxa including mysidae, whereas specificity ranged from 0% for *Paraeuchaeta* sp. females with eggs (only one false positive identification was found) and 99% for copepod eggs. At station 126, accuracy at station 126 ranged from 50% for three groups including gammaridae and 100% for several taxa including *C. hyperboreus* males, whereas specificity ranged from 0% for four taxa including *Paraeuchaeta* sp. C2 and 100% for *Heterorhabdus* sp. Note that sample sizes were very low (i.e. below $n = 5$) for taxa with low accuracy or specificity (Table 4).

The confusion matrices for stations 101 and 126, based on the manual verifications of automatic identifications, showed pattern similar to model internal testing (Figs. 9, 10). Mean probability of misclassifications for stations 101 and 126 was 3.1% and 5.2% respectively. Certain stages of *C. hyperboreus* were wrongly identified as *C. glacialis stages* and vice versa, but at low probabilities. Several misclassifications occurred at higher probabilities at station 126 (e.g. between ostracods and cirripedia cypris stage or between *Oikopleura* sp. and *A. digitale*, Fig. 10).

Counts from automatic identifications at stations 101 and 126 were very similar to counts from the biological sample for most stages of the large copepods *C. hyperboreus*, *C. glacialis* and *M. longa* (e.g. 1109 counts from automatic identifications vs. 1035 ± 123 counts from the biological sample for *C. hyperboreus* C4 at station 1126, Fig. 11). At both stations substantially more individuals of *Pseudocalanus* sp., *Oithona* sp., *Triconia* sp. and *Microcalanus* sp. were identified in the biological sample than by the model. LOKI also encountered difficulties in imaging relatively translucent gelatinous zooplankton. Image data showed that taxa such as *Clione limacina* or chaetognaths, usually abundant in biological samples, were scarce in images.

3.3. Vertical distribution data

In Fig. 12, numbers of *C. glacialis* C5 and *C. glacialis* C3/*C. hyperboreus* C2 counted by the model were used to produce a vertical distribution plot for these taxa.

4. Discussion

This study developed a model to automatically recognize zooplankton specimens in images captured by the Lightframe On-sight Keyspecies Investigation (LOKI) system in the Canadian Arctic. Image quality has only recently increased enough to make such detailed taxonomic classification possible (Schulz et al., 2010; Sainmont et al., 2014). Highly detailed *in-situ* LOKI images allowed us to develop an identification model with 114 original categories, uniquely including many species, and stages and orientations of copepod species. Through three separate validation approaches: (1) model internal testing, (2) comparisons of model results with those of manual identifications of images, and (3) comparisons of model counts with zooplankton numbers actually in the biological samples taken by LOKI, we showed that our model was highly efficient in zooplankton identification, especially for certain taxa. Results of internal model testing (high identification accuracies and specificities of the $n = 114$ classifier as well as of the compressed results with 63 categories) certainly substantiate that claim. Although some categories underperformed, overall model performance is in the range of what an expert labeller can achieve consistently using microscopes (Culverhouse et al., 2003). Consistency is a considerable problem when different taxonomists manually count zooplankton under the microscope (Culverhouse et al., 2014). Accuracy and specificity for many taxa, including stages of large copepods (*C. hyperboreus*, *C. glacialis* and *M. longa*), were good and allow data derived from the automatic identification model to answer ecological questions for these taxa. However, accuracy and specificity was lower for small, translucent, similar looking young stages of mesozooplankton (e.g., *M. longa* C1 & C2, *Oithona* sp., *Microcalanus* sp., *Pseudocalanus* sp. C1). For categories with lower accuracy and specificity the approach described in Solow et al. (2001) can be applied to correct abundance estimates based on the determined misclassification probabilities. The approach makes abundances derived from these categories more reliable.

Validation of the model based on manual identifications of zooplankton in the images for the two completely identified test stations generally showed high agreement with results of model internal testing (at station 101, accuracy and specificity differed by just 1% from internal testing). At station 126, tests showed a 5% difference in accuracy and specificity (this test was completely independent since no images from this station were used for training, therefore these results are particularly promising). The confusion matrix for station 126 showed that some taxa (e.g. Cirripedia cypris stage or Gammaridae) had higher misclassification probabilities (maximum = 50%), however these were typically categories where images and automatic identifications were very rare (sometimes as few as $n = 1$) and therefore misclassification measures were not reliable in those cases. Although the
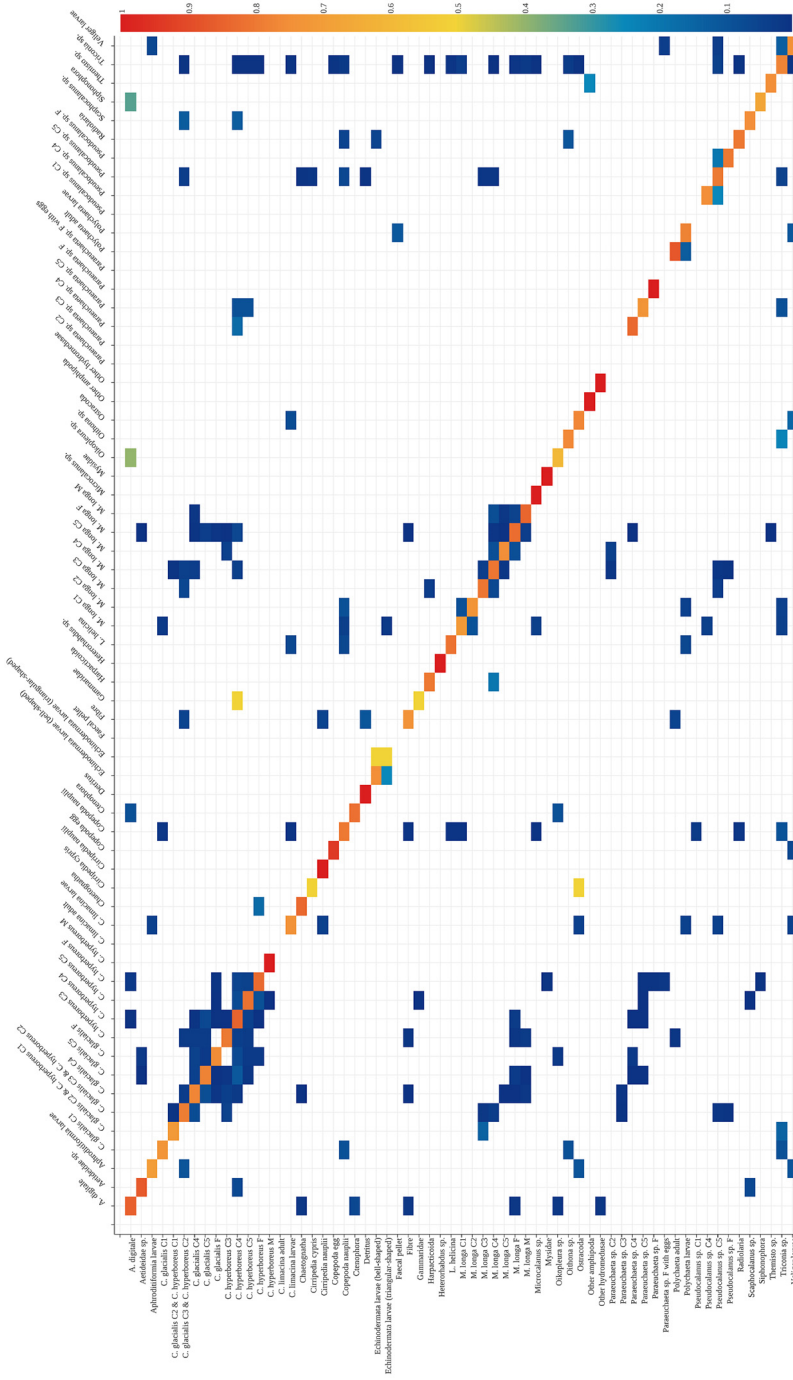
Fig. 9. External testing results for station 101. The confusion matrix shows probabilities of classifications for the 63 compressed categories ($n = 4820$). See the caption of Fig. 7 for more details.

Fig. 10. External testing results for station 126. The confusion matrix shows probabilities of classifications for the 63 compressed categories ($n = 2163$). See the caption of Fig. 7 for more details.
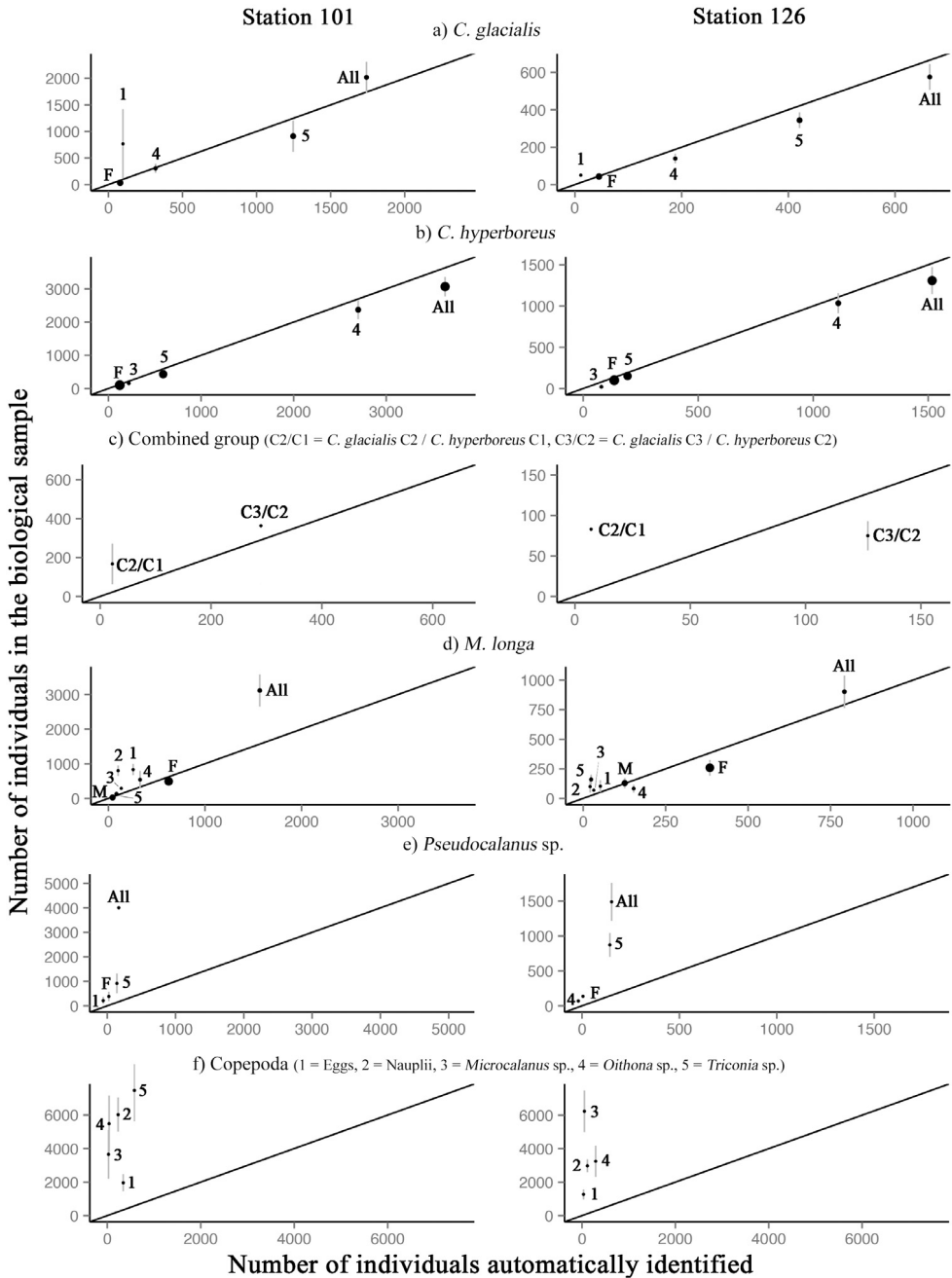
Fig. 11. Numbers of zooplankton specimens in particular groups identified by the model, compared to those present in the biological samples from stations 101 and 126. Numbers 1–5 show copepodite stages: M = males, F = females, All = sum of all stages in the biological sample. The black line indicates perfect agreement.

confusion matrices for stations 101 and 126 were based on 50% randomly labelled ground-truthing datasets we believe the sample size was generally high enough (lower limit of $n = 15$ if possible, leading to 100% labelled images per category in some cases) and that no substantial bias was introduced
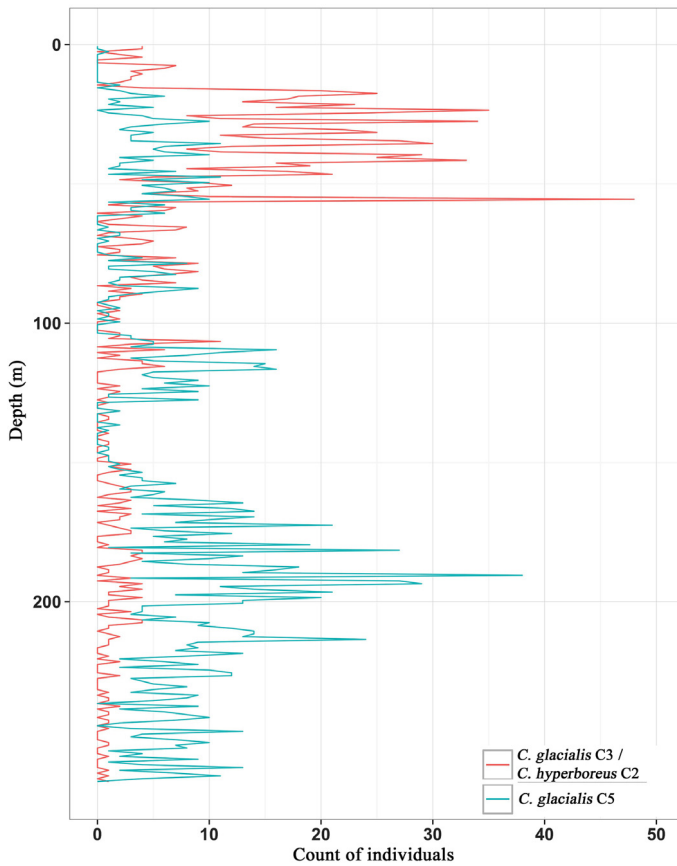
Fig. 12. *C. glacialis* C5 and *C. glacialis* C3/*C. hyperboreus* C2 binned at 1 m intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to accuracy and specificity measures. In comparison to the subsampling employed in the analysis of biological samples where external factors (e.g. gravity) can lead to a bias (e.g. oversampling certain size classes), random subsampling from a list of images must be much less likely to introduce bias.

In our third test, results of the model were compared with actual zooplankton counts in the biological samples. This is a useful test of the model, however note that it assumes that the LOKI camera successfully imaged all animals that passed through during deployment. Furthermore, since only fractions of the biological samples were analysed (total abundances were estimates), whereas the model identified every imaged plankton, counts in the two approaches could be expected to differ. However, for relatively large animals with high sample size (e.g., *C. hyperboreus* and *C. glacialis*), abundances were in fact very similar between the model identifications and the biological sample. Furthermore, for these taxa, results of automatic identifications were similar to the biological sample counts at both test stations, showing consistency with the comparison of automatic identifications and manual identifications of the underlying images. Both external tests therefore point to a good ability of the model to generalize on new data. Very small copepods such as *Oithona* sp. were highly abundant in the biological samples, and are of high importance in the ecosystem (e.g., Gallienne and Robins, 2001 and Turner, 2004). It seems that the LOKI system had problems in imaging such species. Very small, translucent copepods were probably missed because of the algorithm which detects zooplankton when flowing through the camera channel during deployment. If that algorithm is too sensitive LOKI tries to record every particle, whereas if the algorithm is not sensitive enough it misses very small and translucent organisms. Trying to capture pictures of all particles leads to a crash of the camera

framerate and an overflow of the image buffer of the system, therefore rendering data collected neither quantitative nor representative. Therefore that algorithm needs further advancements in order to efficiently sample animals such as *Oithona* sp. or *Triconia* sp.

The prosome length and width models which were developed in this study provided meaningful predictors for the automatic identification model. In copepod taxonomy prosome length rather than total length is often used to identify species and stages (prosome length values can be more specific to a given stage). For copepods, prosome measurements are therefore likely to be better predictors than total length values derived from Equivalent Spherical Diameter (ESD), which are generally used in other zooplankton automatic identification studies (e.g. Bell and Hopcroft, 2008 and Gorsky et al., 2010). Grossmann and Lindsay (2014) developed regression models between ESD and prosome length which can be used to approximate prosome length after identification, obtaining R^2 values from 0.74 to 0.91 for different taxa. The machine learning regressions developed here had higher R^2 values of 0.97 for length and 0.96 for width, over all taxa. Prosome length can also be converted to biomass of each imaged zooplankton using taxa specific conversion factors (e.g. Blachowiak-Samolyk et al., 2008 for the Arctic).

Older studies using automatic zooplankton recognition models generally worked with a much coarser taxonomic resolution than presented here (typically up to 20 categories), and were often based on analyses of preserved specimens (e.g., Hu and Davis, 2006, Bell and Hopcroft, 2008 and Gorsky et al., 2010). More recently, Dieleman et al. (2015) built an automatic identification model for 121 zooplankton categories with ∼82% accuracy using Convolutional Neural Networks, with categories also including shape and orientation subgroups, but again categories had a much coarser taxonomic resolution (e.g., calanoid copepods were not separated into species; see Cowen et al., 2015 for the underlying dataset). Sosik and Olson (2007) developed a successful implementation for automatic identification of 22 phytoplankton categories, solely relying on Matlab for feature extraction and training a Support Vector Machine. Laney and Sosik (2014) then switched to RF for training 47 phytoplankton categories. The RF implementation used in this study is likely superior to other implementations of RF (e.g. in R or Matlab). Reasons may include superior weighting algorithms in Salford's implementation as well as a superior splitting method since Salford's RF is constantly being developed further (Herrick, 2013).

The importance of the different predictors in the classification model showed that the most important predictors belonged to image feature set 1, but also included predicted prosome length and width as well as GUIDOS toolbox's MSPA category branch. Branches characterize the ends of appendages such as antennae, maxillipeds and swimming legs. Adding image feature set 2 to the analysis, which is based on the standalone software GUIDOS toolbox, provided an additional 26 predictors. It added mostly information on the connectivity of pixels in the images (MSPA analysis), and resulted in an increase in overall accuracy and specificity of ∼10%. GUIDOS toolbox was originally developed for landscape ecology and was then repurposed for the method described here. Knowledge transfer, such as then one described here, often lags substantially though (e.g., Cutler et al., 2007). MSPA analysis is only available in GUIDOS toolbox (P. Vogt, personal communication, February 1, 2016) and therefore a pure Matlab approach would miss these important connectivity features in zooplankton images.

The model developed here could be adapted for use in other oceanic regions and with different species by retraining the model. Taxonomic categories and their images can easily be added or taken out from the classification tree developed here. After measuring the parameters of the newly added images, the RF model could rapidly be retrained. In general measuring image parameters is fast. Feature set 1 using LOKI browser is measured fully automatic following the import of station data into the software. Measuring feature set 2 is also fully automatic after setting a few parameters. Several hundred images per minute can be analysed using MSPA analysis and several thousand images per minute using meshsize analysis. "Cleaning" images of artefacts is unfortunately a step that currently has to be done manually to ensure reliability.

4.1. Conclusion

A detailed automatic identification model with 114 original categories was developed for zooplankton taxa imaged with the LOKI and proved via multiple tests to work well, identifying

many species and stages. With this automatic identification model, new images can be identified in a fraction of the time it would take to analyse them traditionally using a binocular microscope, and possibly with higher consistency and accuracy. Comparison of biological samples with automatic identifications revealed that whilst quality of images taken by LOKI is very high, the system did not image all taxa representatively. This is important for the zooplankton optical imaging community since little information exists to date on the comparison of images vs. biological sample. For reliable abundance estimates, it was necessary to remove double images from our dataset. Other researchers using underwater imaging systems should consider whether it is necessary to treat double images in their own approaches (this is not always done). We used the model to produce a high-resolution vertical distribution graph for important zooplankton taxa at 1 m intervals, showing that the model can have powerful applications in studies of zooplankton ecology. The automatic identification model is currently being used to study the coupling between primary and secondary producers in the Arctic.

Improvement of the LOKI imaging system is possible in several ways. The algorithm which cuts out single organism pictures from the overall image frame during LOKI deployment should be improved in order to make the representative imaging of small, but important taxa such as *Oithona* sp. and *Triconia* sp. possible. Colour cameras (Sainmont et al., 2014) are surely going to play a larger role in plankton imaging in the future and are planned for the LOKI system as well.

Acknowledgements

Appendix

See Tables A.1–A.4.

Table A.1
Accuracy (Ac), specificity (Sp) and test sample size for all 114 categories in the automatic zooplankton identification model.

Taxon or particle	Stage	Orientation/ condition	Sample size (*n*)	Ac (%)	Sp (%)
Aetideidae sp.	–	–	64	92	86
Aphroditiformia larvae	–	–	34	83	97
A. digitale	–	Full size	116	90	98
		Partial	67	88	90
Chaetognatha	–	Body	49	93	86
		Head	122	90	75
Cirripedia cypris	–	–	97	85	94
Cirripedia nauplii	–	–	26	94	65
Copepoda egg	–	–	287	100	95
Copepoda nauplii	–	–	362	83	89
Ctenophora	–	Large	39	100	97
		Small	40	97	78

Table A.1 (*continued*)

Taxon or particle	Stage	Orientation/condition	Sample size (*n*)	Ac (%)	Sp (%)
	C1	Dorsal	31	40	71
		Lateral	16	21	63
	C4	Dorsal	28	70	82
		Lateral	17	38	59
C. glacialis		Dorsal	39	94	79
	C5	Dorsal, antenna extended	8	23	63
		Lateral	88	67	82
		Antenna in front	16	100	88
	F	Dorsal long	23	47	65
		Dorsal short	37	100	68
		Lateral	51	75	78
C. glacialis/C. hyperboreus	*C. glacialis C2/C. hyperboreus C1*	Dorsal	10	47	70
		Lateral	9	100	67
	C. glacialis C3/C. hyperboreus C2	Dorsal	31	82	87
		Lateral	16	77	63
	C3	Dorsal, antenna extended	12	100	92
		Dorsal	26	68	81
		Lateral	14	52	93
	C4	Dorsal	49	95	82
		Dorsal, antenna extended	11	86	55
C. hyperboreus		Lateral	218	73	84
	C5	Dorsal	32	100	91
		Dorsal, antenna extended	53	89	96
		Lateral	154	89	92
	F	Dorsal	48	100	98
		Dorsal, antenna extended	38	97	100
		Lateral	138	86	95
	M	lateral	6	100	83
C. limacina adult	–	–	22	100	100
C. limacina larvae	–	Long	55	96	82
		Round	71	87	82
Detritus	–	–	143	83	71
Echinodermata larvae (bell-shaped)	–	–	50	88	60
Echinodermata larvae (triangular-shaped)	–	–	146	97	88
Faecal pellet	–	–	82	88	91
		Cloudlike shape	323	83	89
Fibre	–	Compact shape	310	91	92
		Elongated	84	91	87
Gammaridae	–	–	39	97	92
Harpacticoida	–	Dorsal	51	100	92
		Lateral	17	100	100
Heterorhabdus sp.	–	–	40	91	100

(continued on next page)

Table A.1 (*continued*)

Taxon or particle	Stage	Orientation/ condition	Sample size (*n*)	Ac (%)	Sp (%)
L. helicina	–	–	126	87	92
Microcalanus sp.	–	Dorsal	35	63	69
		Lateral	33	100	42
Mysidae	–	Extended	97	98	98
		Round	40	95	100
	C1	Dorsal	24	94	71
		Lateral	31	55	84
	C2	Dorsal	15	48	73
		Lateral	17	75	88
	C3	dorsal	15	48	73
		Lateral	13	54	54
M. longa	C4	Dorsal	29	79	90
		Lateral	26	69	69
	C5	Dorsal	31	89	81
		Dorsal, antenna extended	17	65	88
		Lateral	33	59	79
	F	Dorsal	44	97	86
		Dorsal, antenna extended	22	78	64
		Lateral	218	84	89
	M	Dorsal	12	100	83
		Dorsal, antenna extended	20	100	65
		Lateral	57	80	70
Oikopleura sp.	–	Complete	36	96	75
		Partial	66	79	83
Oithona sp.	–	–	52	89	62
Ostracoda	–	–	295	82	84
Other amphipoda	–	Dorsal	16	86	75
		Lateral	38	95	95
		Round	67	98	94
Other hydromedusae	–	–	17	100	88
	C2	Lateral	17	100	47
	C3	Lateral	20	100	40
	C4	Dorsal	27	100	78
		Lateral	14	100	71
Paraeuchaeta sp.	C5	Dorsal	21	100	76
		Dorsal, antenna extended	17	100	76
		Lateral	38	97	84
	F	Dorsal	13	86	92
		Lateral	29	100	69
		With eggs	39	95	97
Polychaeta adult	–	Long	84	100	85
		Round	38	76	76
Polychaeta larvae	–	Bulky shape	61	98	84

Table A.1 (*continued*)

Taxon or particle	Stage	Orientation/condition	Sample size (*n*)	Ac (%)	Sp (%)
		Thin shape	94	85	78
	C1	Lateral	20	100	10
	C4	Lateral	10	100	70
Pseudocalanus sp.	C5	Dorsal	61	88	72
		Lateral	159	83	86
	F	Dorsal	24	85	71
		Lateral	50	76	74
Radiolaria	–	–	234	85	91
Scaphocalanus sp.	–	–	39	90	90
Siphonophora	–	–	75	98	69
		Dorsal	38	97	87
Themisto sp.	–	Lateral	75	91	68
		Round	42	95	98
		Dorsal	34	96	74
Triconia sp.	–	Lateral	296	88	91
		M and F attached	99	88	92
Veliger larvae	–	–	41	83	93

Table A.2
Accuracy (Ac) and specificity (Sp) results for the 63 compressed categories.

Taxon or particle	Stage	Sample size (*n*)	Ac (%)	Sp (%)
(a) Copepods identified to the stage level				
	C1	47	68	82
C. glacialis	C4	45	75	81
	C5	135	77	85
	F	127	76	78
	C3	52	75	88
	C4	278	79	87
C. hyperboreus	C5	239	90	92
	F	224	91	96
	M	6	100	83
C. glacialis C2/*C. hyperboreus* C1	–	19	62	72
C. glacialis C3/*C. hyperboreus* C2	–	47	80	79
	C1	55	69	87
	C2	32	67	85
	C3	28	72	72
M. longa	C4	55	76	80
	C5	81	77	83
	F	284	85	88
	M	89	86	77

Table A.2 (*continued*)

Taxon or particle	Stage	Sample size (*n*)	Ac (%)	Sp (%)
Paraeuchaeta sp.	C2	17	100	47
	C3	20	100	42
	C4	41	100	76
	C5	76	98	80
	F	42	94	76
	F with egg	39	95	97
Pseudocalanus sp.	C1	20	100	20
	C4	10	100	70
	C5	220	84	83
	F	74	78	75
(b) Copepods not identified to the stage level				
Aetideidae	–	64	92	89
Copepoda egg	–	287	100	95
Copepoda nauplii	–	362	83	90
Harpacticoida	–	68	100	94
Heterorhabdus sp.	–	40	91	100
Microcalanus sp.	–	68	73	62
Oithona sp.	–	52	89	64
Scaphocalanus sp.	–	39	90	90
Triconia sp.	–	429	88	91
(c) Gelatinous zooplankton				
Aglantha digitale	–	183	90	95
C. limacina adult	–	22	100	100
C. limacina larvae	–	126	90	82
Chaetognatha	–	171	91	79
Ctenophora	–	79	99	87
Limacina helicina	–	126	87	92
Oikopleura sp.	–	102	84	81
Other hydromedusae	–	17	100	88
Siphonophora	–	75	98	69
Veliger larvae	–	41	83	93
(d) Non-gelatinous-non-copepod zooplankton				
Aphroditiformia larvae	–	34	83	97
Cirripedia cypris	–	97	85	94
Cirripedia nauplii	–	26	94	65
Echinodermata larvae (bell-shaped)	–	50	97	60
Echinodermata larvae (triangular-shaped)	–	146	89	88
Gammaridae	–	39	97	92
Mysidae	–	137	97	99
Ostracoda	–	295	82	84
Other amphipoda	–	121	96	92
Polychaeta adult	–	122	76	57
Polychaeta larvae	–	155	90	83
Radiolaria	–	234	85	92
Themisto sp.	–	155	94	83
(e) Other particles				
Detritus	–	143	83	71
Faecal pellet	–	82	88	91
Fibre	–	717	88	91

Table A.3

Predictors ordered according to their overall importance for the three Random Forests models built in this study: (a) the prosome length model, (b) the prosome width model, (c) the automatic zooplankton identification model. Set 1 denotes a feature from LOKI browser and Set 2 originates from GUIDOS toolbox.

(a)

Variable	Score
Set 1, Area in mm²	100
Set 1, Fourier descriptor 1	66.03
Set 1, Fourier descriptor 2	34.64
Set 1, Homogeneity	24.74
Set 2, Meshsize	12.11
Set 1, Hu moment 2	11.91
Set 2, Computation time meshsize	9.98
Set 1, Hu moment 1	7.53
Set 2, Core foreground	4.15
Set 1, Fourier descriptor 4	3.26
Set 1, Skewness	2.98
Set 1, Hu moment 3	2.72
Set 1, Fourier descriptor 6	2.7
Set 1, Convexity	2.54
Set 2, Branch data pixels	2.37
Set 2, Core data pixels	2.23
Set 2, Branch foreground	2.21
Set 1, Fourier descriptor 3	2.2
Set 1, Hu moment 6	2.12
Set 2, Background data pixels	2.05
Set 1, Greymean	1.97
Set 2, Relative fragmentation	1.91
Set 2, Branch frequency	1.88
Set 2, Islet frequency	1.79
Set 1, Fourier descriptor 7	1.76
Set 1, Hu moment 5	1.68
Set 1, Edge frequency	1.65
Set 1, Kurtosis	1.65
Set 1, Form	1.63
Set 1, Fourier descriptor 8	1.57
Set 1, Hu moment 4	1.5
Set 2, Background frequency	1.49
Set 1, Fourier descriptor 9	1.43
Set 1, Hu moment 7	1.41
Set 2, Loop frequency	1.38
Set 2, Core frequency	1.28
Set 2, Perforation frequency	1.28

(b)

Variable	Score
Set 1, Area in mm2	100
Set 2, Meshsize	83.09
Set 1, Fourier descriptor 1	41.32
Set 1, Homogeneity	21.56
Set 1, Fourier descriptor 2	20.81
Set 1, Hu moment 1	19.85
Set 2, Core foreground	19.6
Set 1, Hu moment 2	16.98
Set 2, Computation time meshsize	13.5
Set 2, Relative fragmentation	12.26
Set 2, Background data pixels	11.45
Set 2, Islet frequency	9.41
Set 2, Branch frequency	9.17
Set 1, Convexity	8.84
Set 2, Core data pixels	8.29
Set 1, Skewness	7.86
Set 2, Edge foreground	7.8
Set 1, Form	6.22
Set 2, Edge frequency	6
Set 1, Greymean	6
Set 1, Hu moment 6	5.47
Set 2, Bridge frequency	5.39
Set 2, Perforation frequency	5.08
Set 2, Islet data pixels	5.01
Set 2, Background frequency	4.85
Set 1, Hu moment 4	4.84
Set 1, Fourier descriptor 6	4.56
Set 1, Fourier descriptor 3	4.55
Set 1, Kurtosis	4.47
Set 1, Hu moment 7	4.36
Set 2, Core frequency	4.33
Set 2, Bridge data pixels	4.04
Set 1, Fourier descriptor 4	3.62
Set 2, Islet foreground	3.58
Set 2, Edge data pixels	3.25
Set 1, Hu moment 3	3.14
Set 1, Fourier descriptor 8	2.79

(c)

Variable	Score
Set 1, Hu moment 1	100
Set 1, Hu moment 4	77.42
Set 1, Hu moment 2	50
Predicted prosome length (derived)	50
Set 1, Form	50
Predicted prosome width (derived)	50
Set 1, Fourier descriptor 2	34.85
Set 1, Fourier descriptor 1	26.72
Set 1, Homogeneity	25
Set 1, Area in mm2	25
Set 2, Branch foreground	16.41
Set 1, Convexity	13.62
Set 1, Hu moment 3	13.01
Set 1, Skewness	12.5
Set 2, Edge data pixels	12.5
Set 2, Core foreground	12.5
Set 1, Kurtosis	11.89
Set 1, Fourier descriptor 4	10.58
Set 2, Branch data pixels	8.81
Set 1, Fourier descriptor 3	5.69
Set 1, Hu moment 5	3.75
Set 2, Background data pixels	3.67
Set 2, Perforation data pixels	3.61
Set 1, Hu moment 6	3.42
Set 1, Greymean	3.4
Set 2, Relative fragmentation	2.99
Set 2, Meshsize	2.69
Set 2, Edge foreground	2.25
Set 1, Hu moment 7	1.81
Set 2, Core data pixels	1.75
Set 2, Computation time meshsize	1.54
Set 1, Fourier descriptor 5	1.52
Set 1, Fourier descriptor 10	1.17
Set 2, Islet foreground	1.12
Set 2, Bridge data pixels	1.03
Set 2, Islet data pixels	0.9
Set 2, Islet frequency	0.86

Table A.3 (*continued*)

(a) Variable	Score	(b) Variable	Score	(c) Variable	Score
Set 1, Fourier descriptor 5	1.28	Set 1, Fourier descriptor 9	2.57	Set 2, Edge frequency	0.81
Set 1, Fourier descriptor 10	1.27	Set 2, Bridge foreground	2.56	Set 2, Perforation data pixels	0.79
Set 2, Bridge frequency	1.18	Set 2, Loop frequency	2.5	Set 2, Background frequency	0.75
Set 2, Edge data pixels	1.13	Set 1, Fourier descriptor 7	2.44	Set 1, Fourier descriptor 9	0.72
Set 2, Perforation data pixels	1.12	Set 2, Perforation data pixels	2.41	Set 1, Fourier descriptor 6	0.67
Set 2, Perforation foreground	1.1	Set 2, Branch data pixels	2.37	Set 2, Bridge foreground	0.66
Set 2, Loop data pixels	1.04	Set 1, Fourier descriptor 5	2.15	Set 1, Fourier descriptor 8	0.57
Set 2, Bridge data pixels	1.04	Set 1, Hu moment 5	2.13	Set 1, Fourier descriptor 7	0.57
Set 2, Islet foreground	1.03	Set 2, Perforation foreground	2.03	Set 2, Core frequency	0.44
Set 2, Edge foreground	1.02	Set 1, Fourier descriptor 10	1.99	Set 2, Perforation foreground	0.42
Set 2, Loop foreground	0.99	Set 2, Branch foreground	1.97	Set 2, Loop data pixels	0.25
Set 2, Islet data pixels	0.98	Set 2, Loop data pixels	1.88	Set 2, Loop foreground	0.25
Set 2, Bridge foreground	0.96	Set 2, Loop foreground	1.82	Set 2, Loop frequency	0.09
				Set 2, Perforation frequency	0.07
				Set 2, Bridge frequency	0.06

Table A.4
Complete results of the automatic identifications for the two 2 test stations.

Taxon or particle	Stage	Station 101 (n)	Station 126 (n)
Aetideidae	–	8	15
Aphroditiformia larvae	–	11	10
A. digitale	–	56	62
Chaetognatha	–	23	6
Cirripedia cypris	–	62	2
Cirripedia nauplii	–	10	1
Copepoda egg	–	342	32
Copepoda nauplii	–	232	115
Ctenophora	–	15	11
	F	79	45
C. glacialis	C5	1247	421
	C4	318	188
	C1	98	11
C. glacialis C2/C. hyperboreus C1	–	22	7
C. glacialis C3/C. hyperboreus C2	–	290	127
	F	122	135
	M	5	1
C. hyperboreus	C5	591	193
	C4	2697	1109
	C3	220	79
C. limacina adult	–	9	–
C. limacina larvae	–	58	23
Detritus	–	29	2
Echinodermata larvae (bell-shaped)	–	21	4
Echinodermata larvae (triangular-shaped)	–	28	2
Faecal pellet	–	24	–
Fibre	–	16	19
Gammaridae	–	1	2
Harpacticoida	–	7	5
Heterorhabdus sp.	–	5	2
L. helicina	–	79	16
Microcalanus sp.	–	27	1
Mysidae	–	3	4
	F	624	384
	M	43	126
	C5	83	24
M. longa	C4	328	153
	C3	133	32
	C2	100	21
	C1	256	52
Oikopleura sp.	–	18	5
Oithona sp.	–	39	51
Ostracoda	–	39	13
Other amphipoda	–	20	2
Other hydromedusae	–	5	2
	F	2	2
	F with egg	–	13
	C5	14	11
Paraeuchaeta sp.	C4	16	12
	C3	7	–
	C2	1	–
Polychaeta adult	–	32	8
Polychaeta larvae	–	54	9
	F	23	5
Pseudocalanus sp.	C5	141	143

Table A.4 (*continued*)

Taxon or particle	Stage	Station 101 (n)	Station 126 (n)
	C4	1	4
	C1	3	–
Radiolaria	–	54	20
Scaphocalanus sp.	–	1	8
Siphonophora	–	5	6
Themisto sp.	–	20	4
Triconia sp.	–	581	292
Veliger larvae	–	20	31
Total	–	9418	4083

Contributions of Authors:

Cyril Aubry, Louis Fortier and Jordan Grigor contributed to the conception and design of the study as well as to the acquisition of data. Cyril Aubry contributed to the analysis of data. Cyril Aubry, Louis Fortier and Jordan Grigor contributed to interpretation of data. All parties contributed to revising the manuscript critically. Furthermore all parties approve the version submitted here.

Role of the funding source:

The Canada Foundation for Innovation and ArcticNet, a Network of Centres of Excellence of Canada, which funded this project, did not have any influence on the study.

References

Ardyna, M., Babin, M., Gosselin, M., Devred, E., Rainville, L., Tremblay, J.-É., 2014. Recent Arctic Ocean sea ice loss triggers novel fall phytoplankton blooms. Geophys. Res. Lett. 41, 6207–6212. 2014GL061047. http://dx.doi.org/10.1002/2014GL061047.

Arrigo, K.R., van Dijken, G., Pabi, S., 2008. Impact of a shrinking Arctic ice cover on marine primary production. Geophys. Res. Lett. 35, L19603.

Bell, J.L., Hopcroft, R.R., 2008. Assessment of ZooImage as a tool for the classification of zooplankton. J. Plankton Res. 30, 1351–1367. http://dx.doi.org/10.1093/plankt/fbn092.

Benfield, M.C., Davis, C.S., Wiebe, P.H., Gallager, S.M., Gregory Lough, 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. Part II 43, 1925–1945. http://dx.doi.org/10.1016/S0967-0645(96)00044-6.

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. http://dx.doi.org/10.5670/oceanog.2007.63.

Berger, M., 1990. Convexity. Amer. Math. Monthly 97, 650.

Blachowiak-Samolyk, K., Søreide, J.E., Kwasniewski, S., Sundfjord, A., Hop, H., Falk-Petersen, S., Nøst Hegseth, E., 2008. Hydrodynamic control of mesozooplankton abundance and biomass in northern Svalbard waters (79–81°N). Deep Sea Res. Part II 55, 2210–2224.

Bradstreet, M., Cross, W., 1982. Trophic relationships at high arctic ice edges. Arctic 35, 1–12.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. http://dx.doi.org/10.1023/A:1010933404324.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. http://dx.doi.org/10.1007/BF00994018.

Cowen, R.K., Guigand, C.M., 2008. In situ ichthyoplankton imaging system (ISIIS): system design and preliminary results. Limnol. Oceanogr. Methods 6, 126–132. http://dx.doi.org/10.4319/lom.2008.6.126.

Cowen, R.K., Sponaugle, S., Robinson, K.L., Luo, J., Oregon State University; Hatfield Marine Science Center, 2015. PlanktonSet 1.0: Plankton imagery data collected from F.G. Walton Smith in Straits of Florida from 2014-06-03 to 2014-06-06 and used in the 2015 National Data Science Bowl (NODC Accession 0127422). [WWW Document]. NOAA National Centers for Environmental Information. Dataset. URL (accessed 9.12.14).

Culverhouse, P.F., Macleod, N., Williams, R., Benfield, M.C., Lopes, R.M., Picheral, M., 2014. An empirical assessment of the consistency of taxonomic identifications. Mar. Biol. Res. 10, 73–84. http://dx.doi.org/10.1080/17451000.2013.810762.

Culverhouse, P.F., Williams, R., Reguera, B., Herry, V., GonzlezGil, S., 2003. Do experts make mistakes? A comparison of human and machine indentification of dinoflagellates. Mar. Ecol. Prog. Ser. 247, 17–25. http://dx.doi.org/10.3354/meps247017.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007. Random forests for classification in ecology. Ecology 88, 2783–2792. http://dx.doi.org/10.1890/07-0539.1.

Davis, C.S., Gallager, S.M., Solow, A., 1992a. Microaggregations of oceanic Plankton observed by towed video microscopy. Science 257, 230–232. http://dx.doi.org/10.1126/science.257.5067.230.

Davis, C.S., Gallager, S.M., Haury, L.R., 1992b. The Video Plankton Recorder (VPR): design and initial results. Arch. Hydrobiol.—Beih. Ergeb. Limnol. 36, 67–81.

Dieleman, S., van den Oord, A., Korshunova, I., Burms, J., Degrave, J., Pigou, L., 2015. Classifying Plankton with deep neural networks [WWW Document]. URL http://benanne.github.io/2015/03/17/plankton.html (accessed 3.17.15).

Falk-Petersen, S., Pavlov, V., Timofeev, S., Sargent, J.R., 2007. Climate variability and possible effects on arctic food chains: The role of Calanus. In: Ørbæk, D.J.B., Kallenborn, D.R., Tombre, D.I., Hegseth, D.E.N., Falk-Petersen, D.S., Hoel, D.A.H. (Eds.), Arctic Alpine Ecosystems and People in a Changing Environment. Springer, Berlin, Heidelberg, pp. 147–166.

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybernet. 36, 193–202. http://dx.doi.org/10.1007/BF00344251.

Gallienne, C.P., Robins, D.B., 2001. Is Oithona the most important copepod in the world's oceans? J. Plankton Res. 23, 1421–1432. http://dx.doi.org/10.1093/plankt/23.12.1421.

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. http://dx.doi.org/10.1093/plankt/fbp124.

Grebmeier, J.M., Overland, J.E., Moore, S.E., Farley, E.V., Carmack, E.C., Cooper, L.W., Frey, K.E., Helle, J.H., McLaughlin, F.A., McNutt, S.L., 2006. A major ecosystem shift in the northern Bering Sea. Science 311, 1461–1464.

Grossmann, M.M., Lindsay, D.J., 2014. Estimation of calanoid copepod prosome length from pixel-based measurements derived from the ZooScan and ZooProcess systems. Bull. Plankton Soc. Japan 61, 91–94.

Guelpen, L., van, Markle, D.F., Duggan, D.J., 1982. An evaluation of accuracy, precision, and speed of several zooplankton subsampling techniques. J. Conseil 40, 226–236. http://dx.doi.org/10.1093/icesjms/40.3.226. Conseil Permanent International pour l'Exploration de la Mer.

Haury, L.R., McGowan, J.A., Wiebe, P.H., 1978. Patterns and processes in the time–space scales of Plankton distributions. In: Steele, J.H. (Ed.), Spatial Pattern in Plankton Communities. In: NATO Conference Series, Springer US, pp. 277–327.

Herrick, K., 2013. Predictive Modeling of Avian Influenza in Wild Birds (Ph.D. thesis), University of Alaska Fairbanks, URL: https://scholarworks.alaska.edu/handle/11122/1837.

Hirche, H.J., Barz, K., Ayon, P., Schulz, J., 2014. High resolution vertical distribution of the copepod Calanus chilensis in relation to the shallow oxygen minimum zone off northern Peru using LOKI, a new plankton imaging system. Deep-Sea Res. I 88, 63–73. http://dx.doi.org/10.1016/j.dsr.2014.03.001.

Hu, M.-K., 1962. Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory 8, 179–187. http://dx.doi.org/10.1109/TIT.1962.1057692.

Hu, Q., Davis, C., 2006. Accurate automatic quantification of taxa-specific plankton abundance using dual classification with correction. Mar. Ecol. Prog. Ser. 306, 51–61. http://dx.doi.org/10.3354/meps306051.

Jaeger, J.A., 2000. Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. Landsc. Ecol. 15, 115–130.

Jaffe, J., Moore, K., McLean, J., Strand, M., 2001. Underwater optical imaging: Status and prospects. Oceanography 14, 64–75. http://dx.doi.org/10.5670/oceanog.2001.24.

Knauer, U., Meffert, B., 2010. Fast computation of region homogeneity with application in a surveillance task. In: ISPRS Technical Commission V Symposium, Newcastle, UK, pp 337–342.

Kovahi, R., Provost, F., 1998. Glossary of terms. Mach. Learn. 30, 271–274. http://dx.doi.org/10.1023/A:1017181826899.

Laney, S.R., Sosik, H.M., 2014. Phytoplankton assemblage structure in and around a massive under-ice bloom in the Chukchi Sea. Deep Sea Res. Part II 105, 30–41.

Lischka, S., Riebesell, U., 2012. Synergistic effects of ocean acidification and warming on overwintering pteropods in the Arctic. Global Change Biol. 18, 3517–3528.

Ortner, P.B., Cummings, S.R., Aftring, R.P., Edgerton, H.E., 1979. Silhouette photography of oceanic zooplankton. Nature 277, 50–51. http://dx.doi.org/10.1038/277050a0.

Overland, J.E., Wang, M., Walsh, J.E., Stroeve, J.C., 2014. Future Arctic climate changes: Adaptation and mitigation time scales. Earth's Future 2, 68–74.

Richardson, A.J., 2008. In hot water: zooplankton and climate change. ICES J. Mar. Sci. 65, 279–295. http://dx.doi.org/10.1093/icesjms/fsn028.

Rolke, M., Lenz, J., 1984. Size structure analysis of zooplankton samples by means of an automated image analyzing system. J. Plankton Res. 6, 637–645. http://dx.doi.org/10.1093/plankt/6.4.637.

Sainmont, J., Gislason, A., Heuschele, J., Webster, C.N., Sylvander, P., Wang, M., Varpe, O., 2014. Inter- and intra-specific diurnal habitat selection of zooplankton during the spring bloom observed by Video Plankton Recorder. Mar. Biol. 161, 1931–1941. http://dx.doi.org/10.1007/s00227-014-2475-x.

Schmid, M.S., Aubry, C., Grigor, J., Fortier, L., 2015. ZOOMIE v1.0 (Zooplankton Multiple Image Exclusion) [WWW Document]. URL: https://zenodo.org/record/17928, http://dx.doi.org/10.5281/zenodo.17928 (accessed 5.25.15).

Schulz, J., 2013. Geometric optics and strategies for subsea imaging. In: Watson, J., Zielinski, O. (Eds.), Subsea Optics and Imaging. Woodhead Publishing Limited, UK, pp. 243–276.

Schulz, J., Barz, K., Ayon, P., Ludtke, A., Zielinski, O., Mengedoht, D., Hirche, H.-J., 2010. Imaging of plankton specimens with the lightframe on-sight keyspecies investigation (LOKI) system. J. Eur. Opt. Soc.: Rapid Publ. 5, http://dx.doi.org/10.2971/jeos.2010.10017s.

Serreze, M.C., Francis, J.A., 2006. The arctic amplification debate. Clim. Change 76, 241–264.

Soille, P., Vogt, P., 2009. Morphological segmentation of binary patterns. Pattern Recognit. Lett. 30, 456–459. http://dx.doi.org/10.1016/j.patrec.2008.10.015.

Solow, A., Davis, C., Hu, Q., 2001. Estimating the taxonomic composition of a sample when individuals are classified with error. Mar. Ecol. Prog. Ser. 309–311.

Søreide, J.E., Leu, E., Berge, J., Graeve, M., Falk-Petersen, S., 2010. Timing of blooms, algal food quality and Calanus glacialis reproduction and growth in a changing Arctic. Global Change Biol. 16, 3154–3163.

Sosik, H.M., Olson, R.J., 2007. Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry. Limnol. Oceanogr. Methods 5, 204–216.

Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol. Methods 14, 323–348. http://dx.doi.org/10.1037/a0016973.

Turner, J.T., 2004. The importance of small planktonic copepods and their roles in pelagic marine food webs. Zool. Stud. 43, 255–266.

Valiela, I., 1995. Spatial structure: Patchiness. In: Marine Ecological Processes. Springer, New York, pp. 325–354. (Chapter 11).

Vogt, P., 2014. GuidosToolbox (Graphical User Interface for the Description of image Objects and their Shapes) [WWW Document]. URL: http://forest.jrc.ec.europa.eu/download/software/guidos (accessed 1.3.15).

Von Hippel, P.T., 2005. Mean, median, and skew: Correcting a textbook rule. J. Stat. Educ. 13.

Wassmann, P., Duarte, C.M., Agusti, S., Sejr, M.K., 2011. Footprints of climate change in the Arctic marine ecosystem. Global Change Biol. 17, 1235–1249.

Welch, H., Bergmann, M., Siferd, T., Martin, K., Curtis, M., Crawford, R., Conover, R., Hop, H., 1992. Energy-flow through the marine ecosystem of the lancaster sound region, arctic Canada. Arctic 45, 343–357.

Westfall, P.H., 2014. Kurtosis as Peakedness, 1905–2014. R.I.P. Amer. Statist. 68, 191–195.

Wiebe, P.H., Benfield, M.C., 2003. From the Hensen net toward four-dimensional biological oceanography. Prog. Oceanogr. 56, 7–136. http://dx.doi.org/10.1016/S0079-6611(02)00140-4.

Wojnar, L., Kurzydlowski, K., 2000. Analysis and interpretation. In: Practical Guide to Image Analysis. ASM International, pp. 145–202. (Chapter 7).

Zhang, D., Lu, G., 2002. Shape-based image retrieval using generic Fourier descriptor. Signal Process., Image Commun. 17, 825–848. http://dx.doi.org/10.1016/S0923-5965(02)00084-X.