ARTICLE IN PRESS
Methods in Oceanography ▮ (▮▮▮▮) ▮▮▮–▮▮▮

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 xxxx

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;

Please cite this article in press as: Failletaz, R., et al., Imperfect automatic image classification successfully describes plankton distribution patterns. Methods in Oceanography (2016), http://dx.doi.org/10.1016/j.mio.2016.04.003