Estuarine, Coastal and Shelf Science 162 (2015) 151–160

Contents lists available at ScienceDirect

Estuarine, Coastal and Shelf Science

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

Increasing the quality, comparability and accessibility of phytoplankton species composition time-series data

Adriana Zingone [a, *], Paul J. Harrison [b], Alexandra Kraberg [c], Sirpa Lehtinen [d], Abigail McQuatters-Gollop [e], Todd O'Brien [f], Jun Sun [g], Hans H. Jakobsen [h]

[a] Integrative Marine Ecology Dept, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Napoli, Italy
[b] Dept. Earth & Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[c] Biologische Anstalt Helgoland (AWI), Kurpromenade, 20127498 Helgoland, Germany
[d] Marine Research Centre, Finnish Environment Institute (SYKE), Erik Palménin aukio 1, Helsinki, Finland
[e] SAHFOS, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK
[f] NOAA – NMFS – Science & Technology, Marine Ecosystems Division, 1315 East-West Hwy, Silver Spring, MD 20910, USA
[g] College of Marine Science and Engineering, Tianjin University of Science and Technology, Tianjin 300457, China
[h] Bioscience, Aarhus University, Roskilde 4000, Denmark

ARTICLE INFO

Article history:
Accepted 4 May 2015
Available online 14 May 2015

Keywords:
phytoplankton
Long Term Ecological Research (LTER)
metadata
methodology
quality assurance
quality control
time series

ABSTRACT

Phytoplankton diversity and its variation over an extended time scale can provide answers to a wide range of questions relevant to societal needs. These include human health, the safe and sustained use of marine resources and the ecological status of the marine environment, including long-term changes under the impact of multiple stressors. The analysis of phytoplankton data collected at the same place over time, as well as the comparison among different sampling sites, provide key information for assessing environmental change, and evaluating new actions that must be made to reduce human induced pressures on the environment. To achieve these aims, phytoplankton data may be used several decades later by users that have not participated in their production, including automatic data retrieval and analysis. The methods used in phytoplankton species analysis vary widely among research and monitoring groups, while quality control procedures have not been implemented in most cases. Here we highlight some of the main differences in the sampling and analytical procedures applied to phytoplankton analysis and identify critical steps that are required to improve the quality and intercomparability of data obtained at different sites and/or times. Harmonization of methods may not be a realistic goal, considering the wide range of purposes of phytoplankton time-series data collection. However, we propose that more consistent and detailed metadata and complementary information be recorded and made available along with phytoplankton time-series datasets, including description of the procedures and elements allowing for a quality control of the data. To keep up with the progress in taxonomic research, there is a need for continued training of taxonomists, and for supporting and complementing existing web resources, in order to allow a constant upgrade of knowledge in phytoplankton classification and identification. Efforts towards the improvement of metadata recording, data annotation and quality control procedures will ensure the internal consistency of phytoplankton time series and facilitate their comparability and accessibility, thus strongly increasing the value of the precious information they provide. Ultimately, the sharing of quality controlled data will allow one to recoup the high cost of obtaining the data through the multiple use of the time-series data in various projects over many decades.

1. Introduction

Phytoplankton time-series (PTS) data are important to assess the ecological health and status of water bodies and changes occurring under climatic and anthropogenic pressures. When

* Corresponding author.
E-mail address: zingone@szn.it (A. Zingone).

http://dx.doi.org/10.1016/j.ecss.2015.05.024
0272-7714/

supplemented with a suite of physical (e.g. temperature, salinity), chemical (nutrients), and biological variables (e.g. chlorophyll and zooplankton data), PTS can be used to evaluate long-term changes in pelagic systems and possible causes and consequences of changes on the marine ecosystem. Marked interannual differences in the response of phytoplankton to seasonal forcing factors are a common characteristic observed among and within coastal sites (Cloern and Jassby, 2010; Zingone et al., 2010). Therefore, time series over several decades are required to discern statistically significant climate-driven trends from random events (Henson et al., 2010). However, multi-decadal PTS, particularly those containing species composition data, are relatively rare (Edwards et al., 2010).

For many relevant issues such as biodiversity, Harmful Algal Blooms (HABs), food-web structure, the invasion of non-indigenous/alien species and ecological process studies, taxonomic information on plankton composition at the species level is required. In the case of HAB species, correct taxonomic identification is needed to inform decision-making in order to alert the public and possibly close areas to shellfish harvesting. Indeed, phytoplankton species composition is part of a suite of ecological indicators that are required to assess the ecological/environmental status of a water body as mandated by the EU Water Framework Directive 2000/60/EC (European Commission, 2000), Marine Strategy Framework Directive 2008/56/EC (European Commission, 2008), and other regional regulation, such as the State Oceanic Administration of China.

Presently, microscope-based identification and enumeration is the 'gold standard' to which other approaches to determine phytoplankton composition (e.g. HPLC, image analysis, flow cytometry) are compared. However, analysis by light microscopy (LM) is time consuming and difficult to standardise, and generally requires highly specialised taxonomic expertise. In addition, it may have limited resolution power for certain groups of phytoplankton species, such as the small flagellates. To study these taxa, molecular methods have been introduced in some time-series studies (e.g., Medlin et al., 2006; McDonald et al., 2007), also helping identify morphologically similar/identical species (cryptic species) which may have distinct ecological, biogeographic and phenological patterns (Degerlund et al., 2012; Ruggiero et al., 2015). Recent DNA metabarcoding approaches using high throughput sequencing (HTS) techniques represent an impressive advancement (e.g. Massana et al., 2002; Guillou et al., 2004). Rapid technological developments are expected to allow routine identification based on environmental DNA analyses, which will possibly be incorporated into automated detection systems in the future. Currently these methods are not quantitative, as they provide relative abundances that may be biased due to preferential sequence amplification, nor are they exhaustive, as reference molecular information is still lacking for most phytoplankton species. Verification and inter-calibration with microscope-based identification, and a coupling of molecular and morphological approach is recommended (McManus and Katz, 2009). Molecular methods are also diverse and have complex metadata for which a level of standardization comparable to other types of data does not currently exist (Sansone et al., 2012). This presents a considerable challenge and for the foreseeable future molecular surveys cannot entirely replace traditional taxonomic surveys.

Semi-automated image analysis of samples by a laboratory-based FlowCAM (Sieracki et al., 1998; Jakobsen and Carstensen, 2011), fluorescent image analysis with PlanktoVision (Schulze et al., 2013), submersed flow-cytometers such as the CytoBuoy (Dubelaar et al., 1999), the CytoSense scanning flow-cytometer (Malkassian et al., 2011) and the flow Cytobot with imaging options (e.g., Olson and Sosik, 2007) show promise for automatic classification of phytoplankton. These instruments can generate thousands of images per hour which preclude manual inspection to verify cell identification. Hence, new challenges in this field are in the development of techniques to analyse these large datasets of phytoplankton images (e.g., Álvarez et al., 2012), while expert taxonomists are required to 'train' automated systems for taxa recognition. On the other hand, these methods are consistent in the generation of their errors and do often provide precise information on the magnitude of their errors (e.g., Culverhouse et al., 2003; Culverhouse, 2007). Their imaging capability is most efficient in analysing individual cells from 10 to 100 μm (Olson and Sosik, 2007) which at times allows one to detect and track taxa of interest with distinct shapes (e.g., Campbell et al., 2010). However, limited image resolution and constraints in the window size observed often do not provide the species level detail that is needed to track changes in HABs, food-webs, and biodiversity in response to changes in climate. Switching from microscopy-based analysis to automated image analysis without considerable inter-calibration of the methods will likely result in inconsistent datasets.

Because species data over time are imperative for understanding changes in the phytoplankton community, most PTS continue to obtain data through LM analyses. Compared to abiotic oceanographic data, these analyses are more complex and difficult to standardise. Differences in sample collection, handling and observation techniques, and the diverse levels of taxonomic experience by the operators, may hamper the reliability of species composition and abundance assessments (Culverhouse et al., 2003; Jakobsen et al., this issue). In turn, the difficulty in controlling and assessing data quality may weaken statistical comparisons of phytoplankton diversity among different PTS datasets. Problems may exist even within a single PTS due to changes in methods, microscopes and analysts over time. For example, it may not always be possible to determine if a new species at a site is an alien/invasive species since it may have been previously misidentified or overlooked by a different person or method.

The development of standards for analysis, taxonomic identification and metadata recording, as well as quality assurance/quality control (QA/QC) procedures, are crucial in order to facilitate a reliable evaluation of the changes in species composition in PTS data, as well as the comparison among PTS collected at different sites. Despite the numerous efforts to standardise PTS data collection, different procedures are still followed to obtain these data. In addition, these procedures are not always reported in sufficient detail in scientific or dataset publications, where there is often a lack of detailed metadata and complementary information.

In an accompanying paper (Harrison et al., this issue), we examined the cell biovolume of 214 ecologically important species in 36 studies and found that incorrect or ambiguous species identification, taxa name changes, and lack of metadata presented limitations to using disparate datasets. That work highlighted the need for the current paper, which addresses metadata recording and QA/QC issues associated with phytoplankton species composition time series. Based on personal experience with time-series data and on information from the literature, we briefly review the different steps of PTS data collection. Our main objective is to identify the key methodological issues which may cause major differences among PTS data from different sites and geographic regions. With the aim of providing a first step towards a comprehensive manual on QA/QC of PTS, we offer some general recommendations which would allow a more reliable comparison among phytoplankton time-series datasets and ensure safe data storage and correct long-term use of these precious data.