
Functional Data Analysis and Classification of Phytoplankton Based on Autonomous Flow Cytometer Data

Anthony Malkassian*^{†1}, David Nérini², Mark A. Van Dijk³, Melilotus Thyssen⁴, Claude Manté⁵, and Gérald Grégori⁶

¹Laboratoire de Microbiologie, de Géochimie et d'Ecologie Marines – CNRS, CNRS – Centre d'Océanologie de Marseille, France

²Laboratoire de Microbiologie, de Géochimie et d'Ecologie Marines – CNRS, CNRS – Centre d'Océanologie de Marseille, France

³Netherlands Institute of Ecology (NIOO-KNAW) – Department of Microbial Ecology, Netherlands

⁴Laboratoire de Microbiologie, de Géochimie et d'Ecologie Marines – CNRS, CNRS – Centre d'Océanologie de Marseille, France

⁵Laboratoire de Microbiologie, de Géochimie et d'Ecologie Marines – CNRS, CNRS – Centre d'Océanologie de Marseille, France

⁶Laboratoire de Microbiologie, de Géochimie et d'Ecologie Marines – CNRS, CNRS – Centre d'Océanologie de Marseille, France

Abstract

Flow cytometry is a useful tool to analyse phytoplankton communities in the ocean. By measuring the light scatter and fluorescence properties on each particle, it provides a fingerprint which enables groups to be separated. A flow cytometer called Cytosense (Cytobuoy-bv, Netherlands - www.cytobuoy.com) has been developed for autonomous in situ sampling and analysis, allowing to follow phytoplankton dynamics at short temporal scale. This instrument is able to record full pulse shape information for each particule on each variable, generating large and complex data sets. Usually, manual analysis is performed and typically consists of visualizing the data set with histograms and two- or three-dimensional scatterplots, and to discriminate groups by drawing regions around them. Our purpose is to provide more objectivity in the data analysis by applying an automatic and consistent method to discriminate groups analysed by the Cytosense. In other words, we seek for partitioning methods on the optical fingerprints of particles, which are complex objects (i.e. 5 curves) by taking into account the shape, the length and the amplitude of each curve. First, numerical experiments were done on simulated profiles. Then the method is applied and validated on mixed phytoplankton cultures, for which clusters overlapped considerably using classical methods of analysis.

Keywords: phytoplankton, automated flow cytometry, multivariate statistics, pulse shape analysis, clustering

*Speaker

[†]Corresponding author: anthony.malkassian@univmed.fr