DESCRIPTION OF THE CLASS

CODE	TITLE MÉTHODS FOR THE EXPLOITATION OF DATA IN OCÉANOGRAPHY
MU5MRM32 **6 ECTS** **M2**	**KEYWORDS:** Data processing and flow; Data analysis: multivariate, spatial, series; Data interpretation **PROFESSORS IN CHARGE:** Jean-Olivier IRISSON, Villefranche sur mer, Sorbonne Université Stéphane GASPARINI, Villefranche sur mer, Sorbonne Université **OTHER PROFESSORS:** Laurent COPPOLA, Villefranche sur mer, Sorbonne Université Christophe MIGON, Villefranche sur mer, Sorbonne Université Laure MOUSSEAU, Villefranche sur mer, Sorbonne Université Lars STEMMANN, Villefranche sur mer, Sorbonne Université

CLASS MODALITIES

TEACHING MODALITIES. 30 hours of lessons, 30 hours of computer labs, some focused on the application of the lessons and others for a tutored project (processing of data collected as part of the UE MU5MRM31 - IADO).

EVALUATION PROCEDURES. The theoretical courses are evaluated by a written exam (2 hours). The project, carried out in pairs, consists of analyzing data from a campaign. It is delivered in the form of a poster and an oral defense (~10 min).

CLASS SUMMARY

The MEDO teaching unit confronts you with the processing of samples and the exploitation of data from a real oceanographic campaign, with their share of imperfections and difficulties. You use laboratory instruments, including plankton imaging ones, and various digital tools, including a programming language, to set up a fast and automated flow between the raw data and your final analyzes, in order to answer a scientific question.

Its principles and organization are described here: https://www.youtube.com/watch?v=tcFrfx-slmc

TEACHING GOALS

At the end of this course, you will know how to:

1. use quantitative imagery to process and taxonomically classify plankton samples;
2. choose and implement statistical techniques for interpolation, multivariate ordination and clustering, and machine learning;
3. program data processing workflows using the R language and visualize data using Ocean Data View software;
4. use physics, biogeochemistry and biology data to answer a specific scientific question;
5. present your results graphically and synthetically.

PRÉREQUISITES

General knowledge of oceanography (physical, biogeochemical and biological), mostly offshore.
Knowledge of classical inferential statistics (variance, ANOVA, regression, correlation).
To make the best use of the data, it is advisable to have followed the EU during which they were collected (IADO – MU5MRM31).

BIBLIOGRAPHY / WEBOGRAPHY

N/A

HOW THE CLASS WORKS

During the first two weeks, you carry out three activities in parallel:

- You process samples collected during the IADO campaign (MU5MRM31). You mainly focus on plankton samples (because the chemical assays were done during the previous course). This involves: (i) measuring chlorophyll content by spectrophotometry, (ii) sorting zooplankton samples under a binocular magnifying glass, (iii) digitizing zooplankton samples using the ZooScan and sorting the images generated on EcoTaxa, using artificial intelligence.

- You learn to use tools to digitally process data: (i) the R programming language, (ii) the Ocean Data View software. The objective is to build, at the end of the two weeks, a completely automated and controlled flow of raw data to their analyzable version.

- You become familiar with various data analysis techniques: signal processing, multivariate ordinations (ACP, AFC), supervised and unsupervised classification, mapping and interpolation, etc. Each technique is seen in class and then practiced during computer labs.

As a result, you have various reformatted and cleaned datasets and the knowledge and tools to analyze them.

The third week of the class is dedicated to a tutored project, carried out in pairs. A list of topics based on the available data is proposed to you; they cover various fields: coastal or offshore oceanography; physics, biogeochemistry, biology or technology; etc. You choose a topic, define a precise question and carry out the analyzes necessary to make the best use of the data and answer it. You finally present your results in the form of a poster and orally.