



HAL
open science

Time series analysis of CTD profiles from the SOMLIT dataset : an approach with functional data analysis

Mathilde Couteyen Carpaye, David Nerini, Gérald Grégori, Team Somlit

► To cite this version:

Mathilde Couteyen Carpaye, David Nerini, Gérald Grégori, Team Somlit. Time series analysis of CTD profiles from the SOMLIT dataset : an approach with functional data analysis. EVOLECO 2024, Nov 2024, Plouzané (Brest), France. hal-04799605

HAL Id: hal-04799605

<https://hal.science/hal-04799605v1>

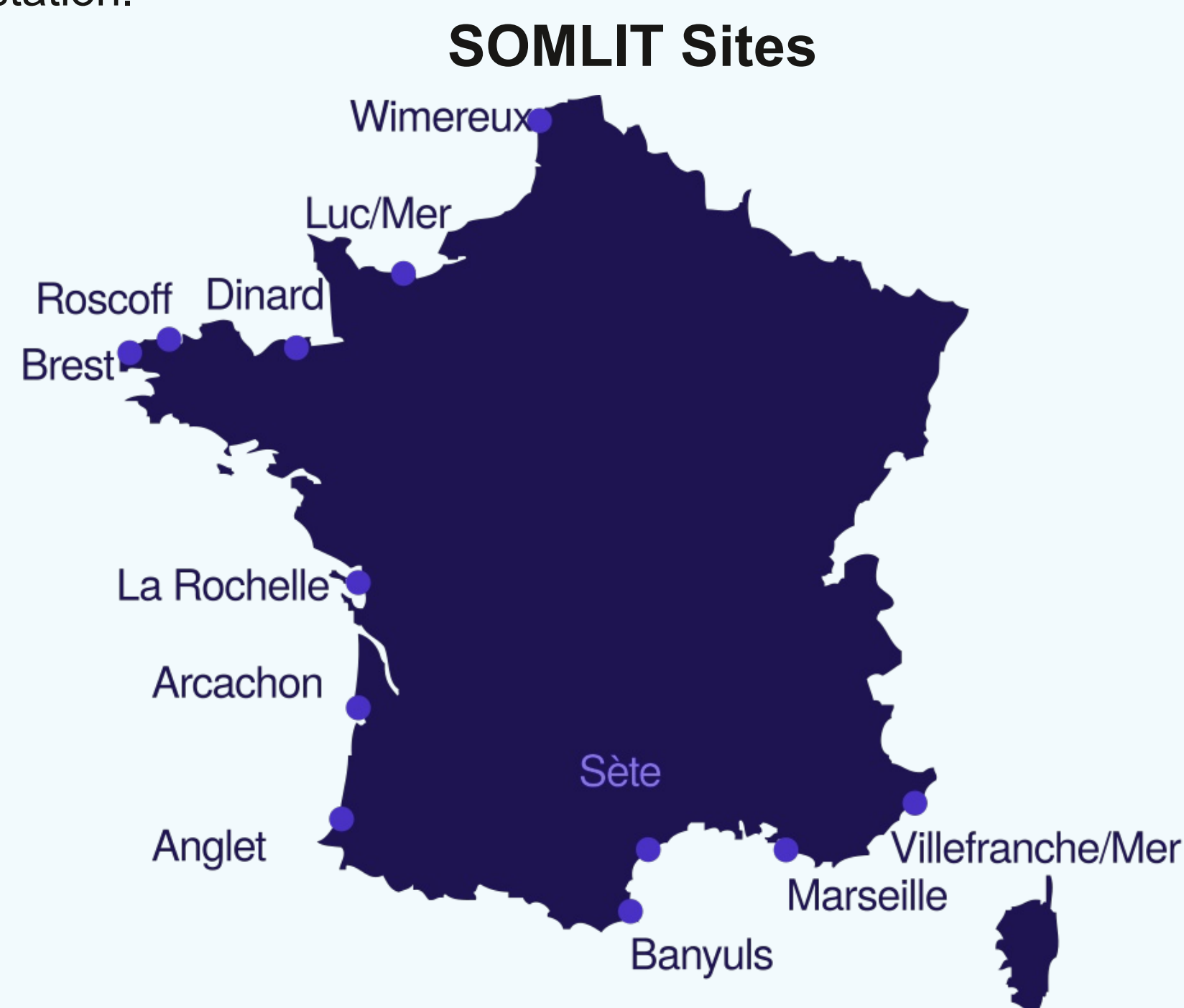
Submitted on 23 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction

Long-term observations are crucial for understanding the sensitivity of microbial assemblages in coastal marine ecosystems to both anthropogenic and natural climate variations, and how local factors can override global changes (Karl & Church, 2014). In France, the National Observation System "SOMLIT" (www.somlit.fr) has been monitoring 12 sites along the metropolitan coast for several decades. As the Mediterranean Sea is considered a hotspot for biodiversity and climate change, we focus on three Mediterranean sites—Banyuls, Marseille, and Villefranche—where data have been collected since 2007. We present a focus on the Banyuls station.



Available data

Surface data

- Physico-chemical variables
- Pico-nanoplankton optic data from flow cytometry

Vertical profiles from CTD* sensors

- Temperature
- Salinity
- Fluorescence

Frequency : bimensual

*Conductivity, Temperature, Depth (CTD) sensors often include additional instruments such as sensors to measure chlorophyll fluorescence

Dataset complexity : vertical, temporal and spatial dimensions

We propose an approach to address the vertical and temporal dimensions of the CTD data.

- Objective: Identify time periods that deviate from the usual observed patterns.

Case study: Anomaly in the spring bloom observed at Banyuls in 2021 and 2022.

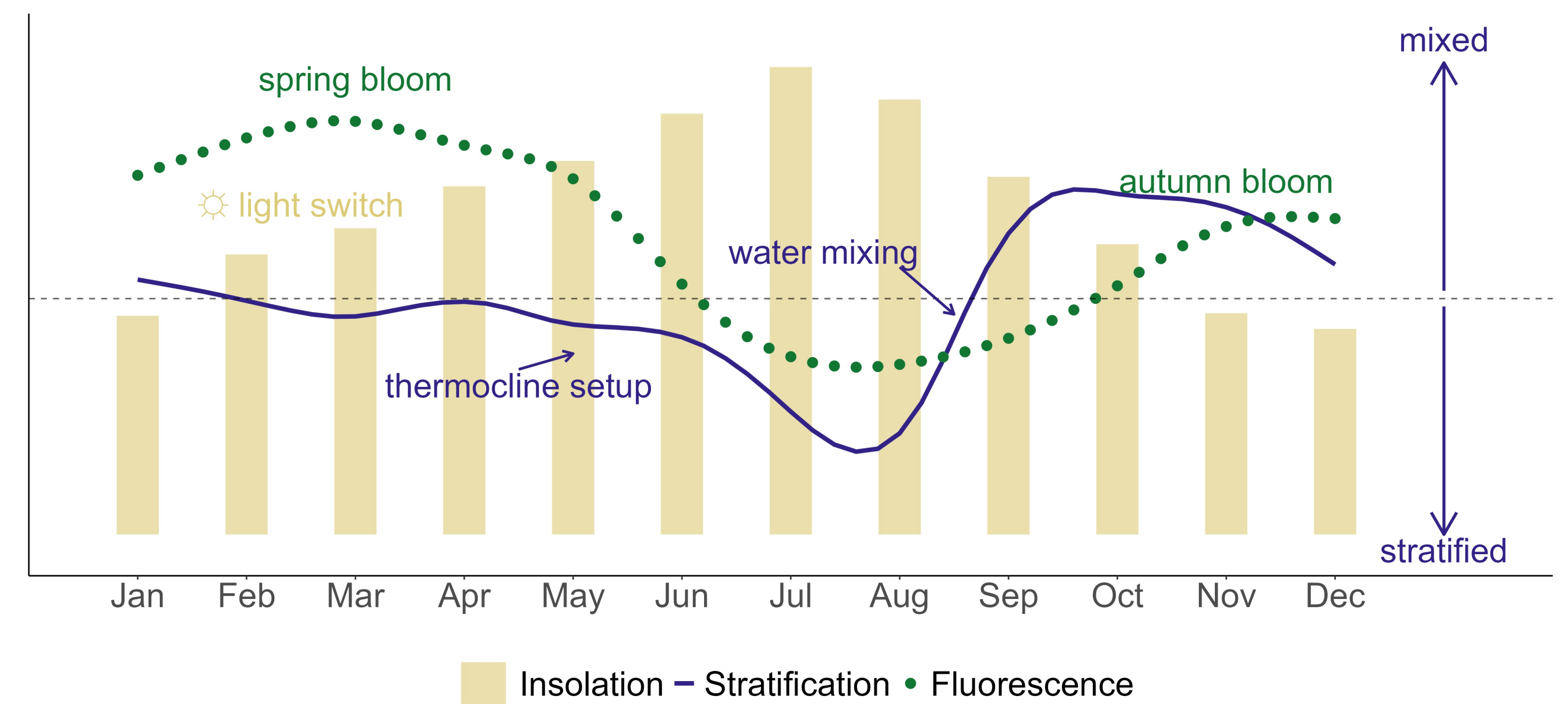
Yearly pattern at Banyuls site

Winter: The water column is mixed. Nutrients are available to phytoplankton at the surface. Limited sunlight prevents rapid growth.

Spring: The water remains mixed. The increase in daily insolation triggers the spring bloom.

Summer: The surface water heats up. Formation of the thermocline: it prevents nutrient supply from the depths to the surface.

Autumn: Increased turbulence at the surface mixes the water column. It enriches the upper layers and triggers the autumn bloom.



Functional Data Analysis (FDA) & Functional Principal Component Analysis (FPCA)

Functional data analysis (FDA) is the branch of statistics that deals with data that arrive as curves. FPCA is a dimension reduction method which allows to study the structure of a curve dataset (Nerini *et al.*, 2022).

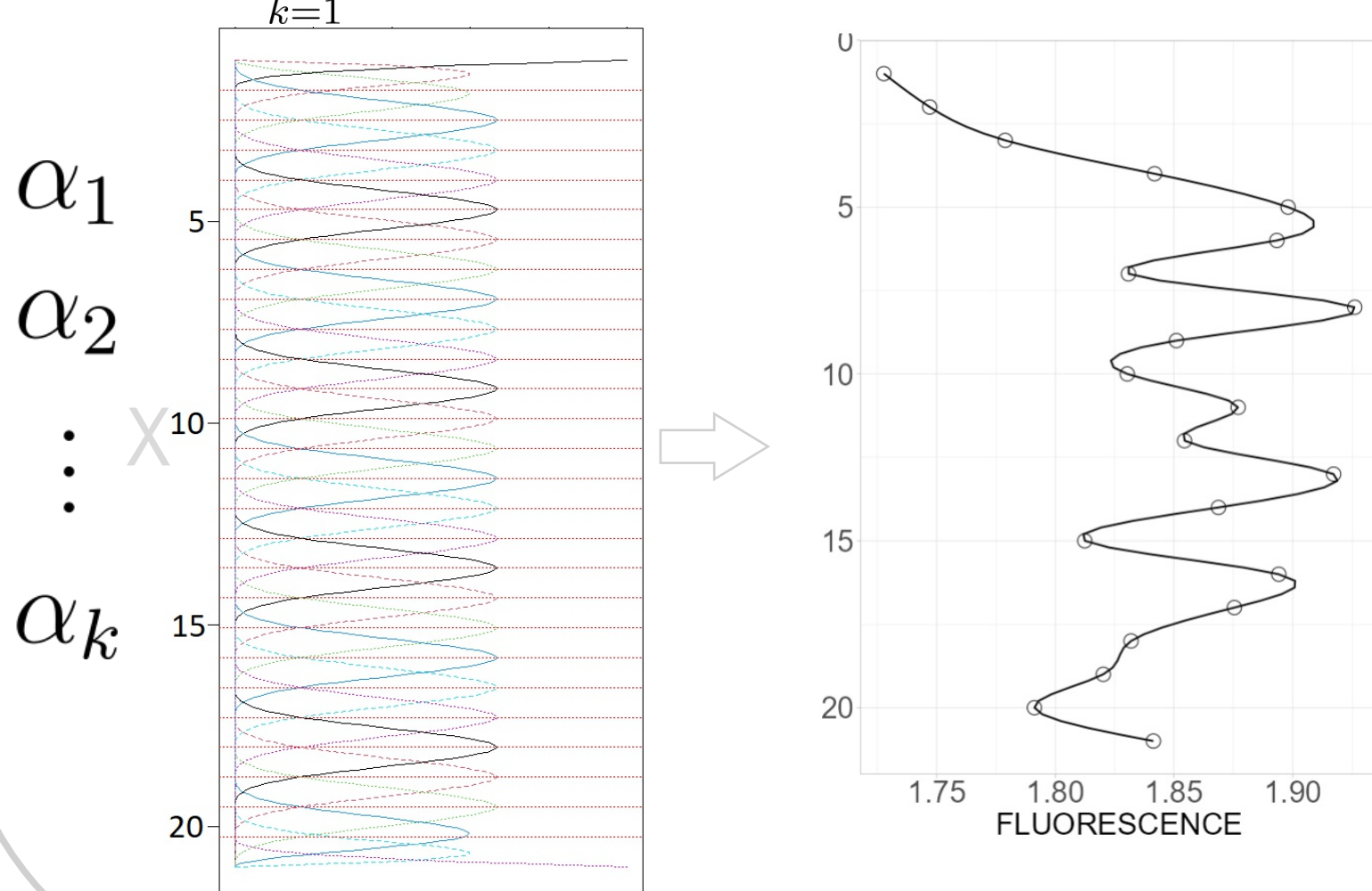
$$\tilde{Z}(t) = \mu(t) + \sum_{q=1}^Q c_q \xi_q(t)$$

continuous process over t → mean curve → eigen functions → principal components (PC) → $\mu(t) \pm \sqrt{\lambda_q} \xi_q(t)$

Methods & Results

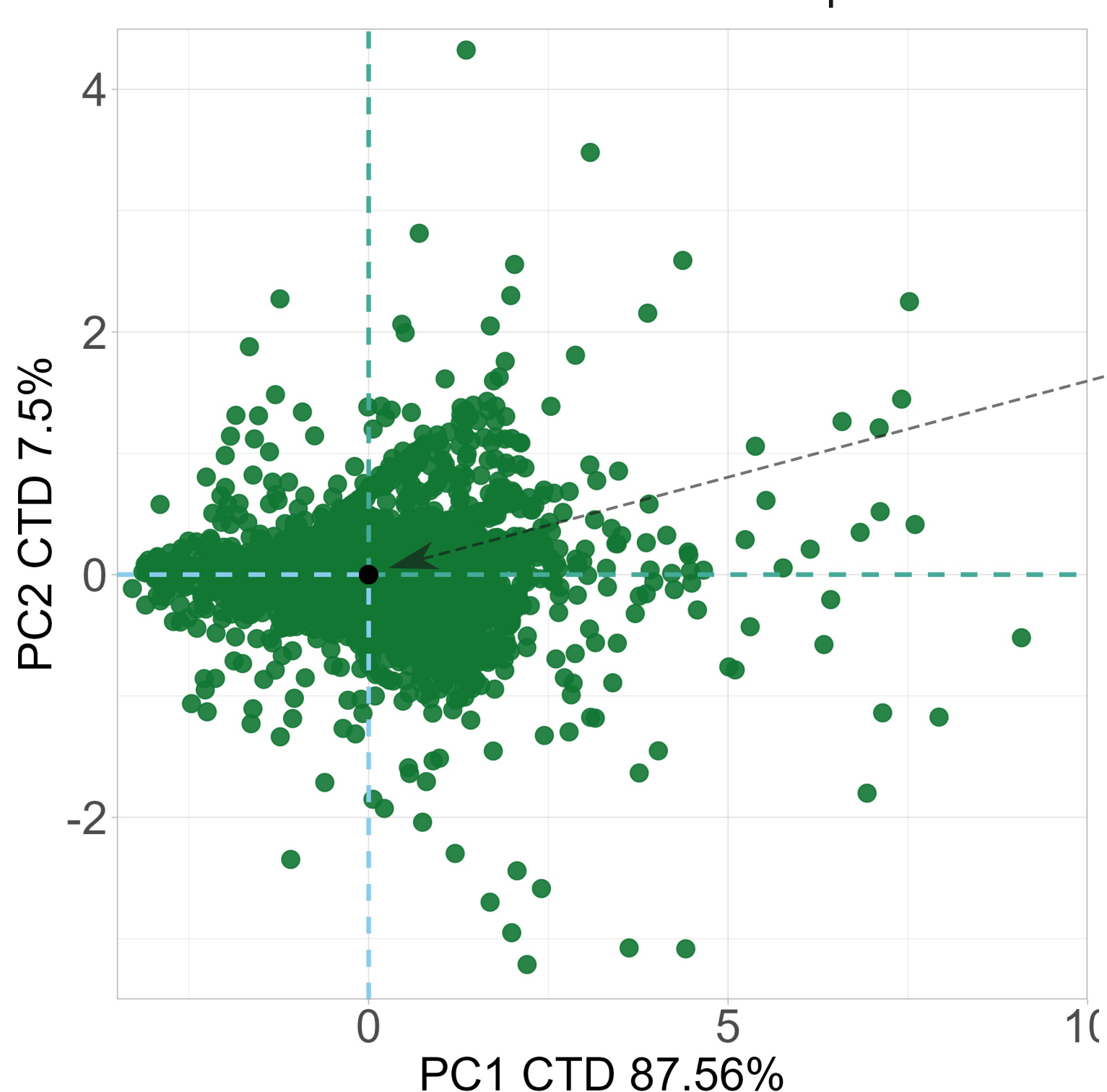
1 Spline smoothing of raw vertical data

$$x_i(t) = \sum_{k=1}^K \alpha_{ik} \phi_k(t), t \in [1m; 21m]$$



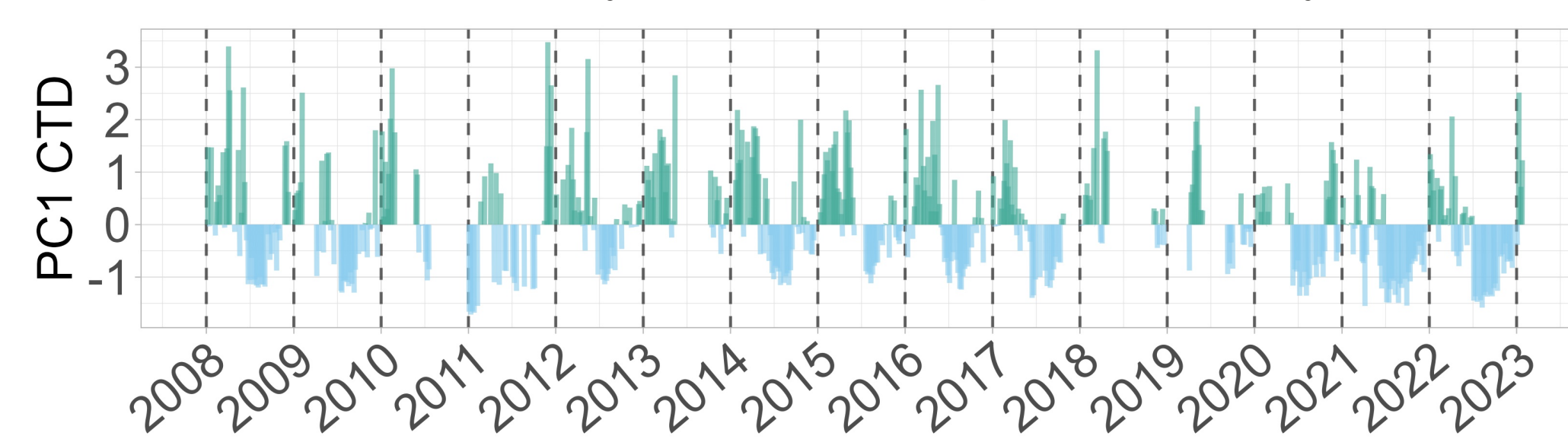
2 FPCA on profiles curves

Dimensional reduction of the smoothed fluorescence vertical profiles.

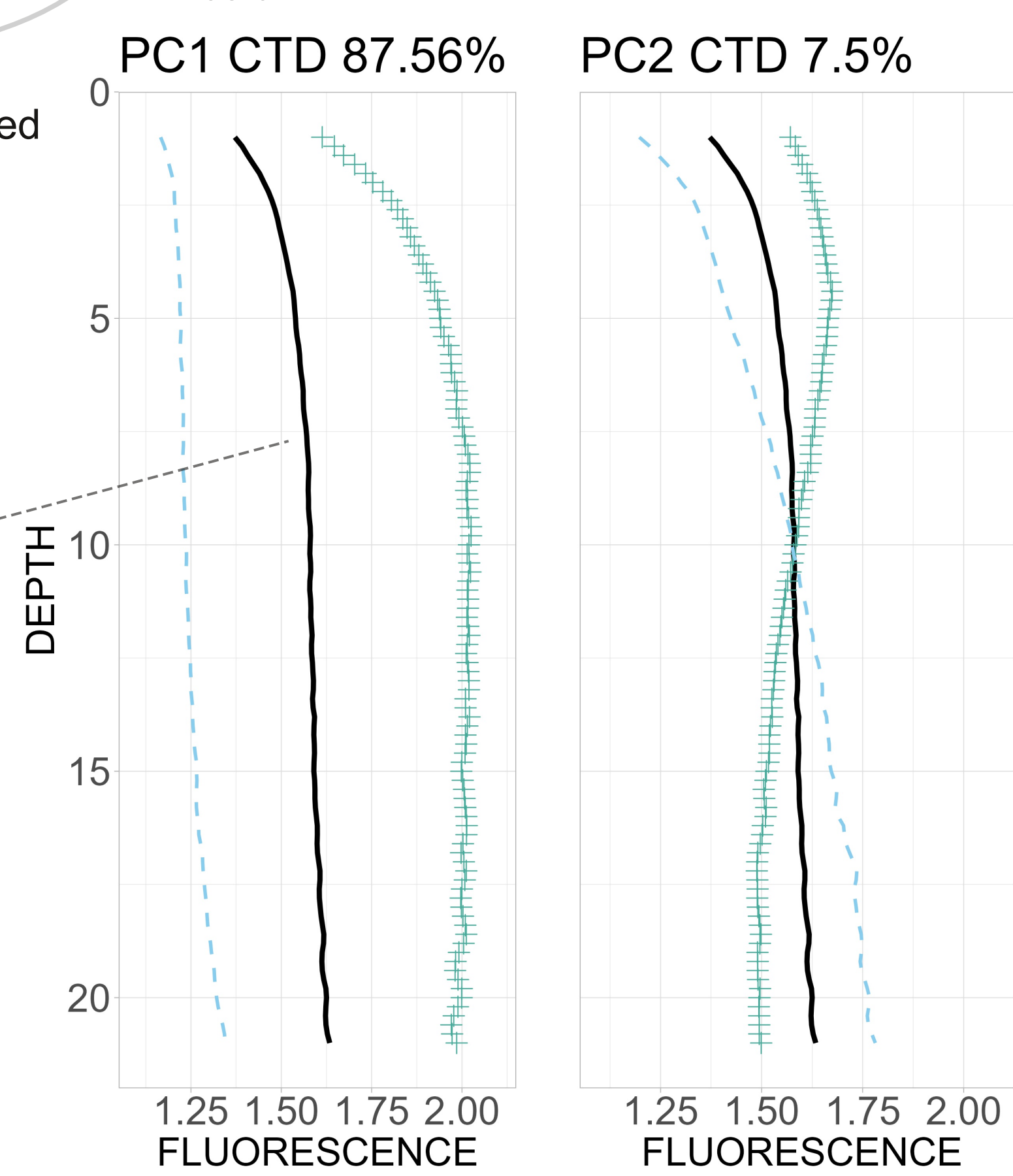


3 Monthly means of the PC1

The scores of each profiles on the first axis are expressed over time. Monthly means are computed for each year.

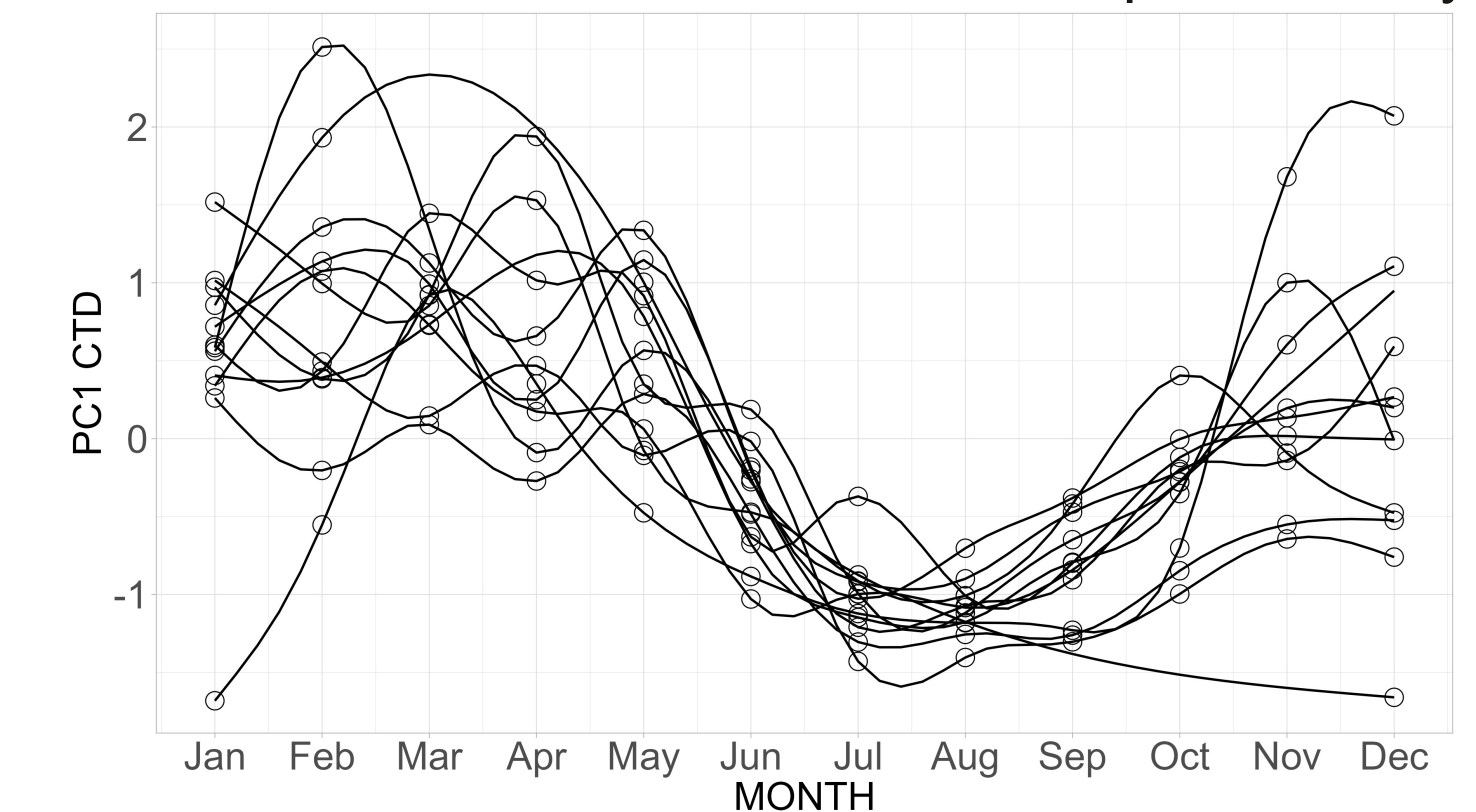


The first principal component depicts an average variation in fluorescence across the entire water column.



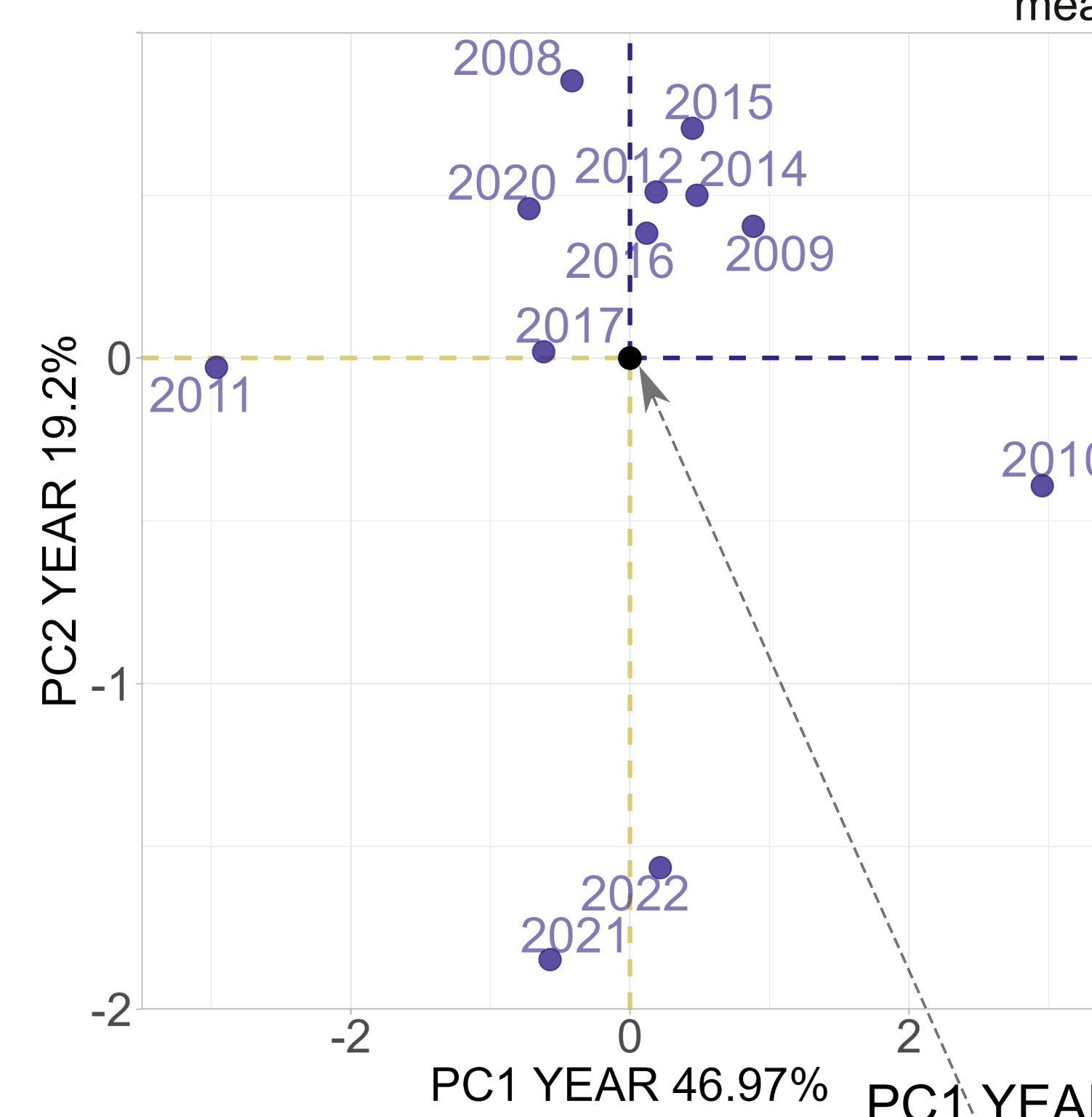
4 Regression spline smoothing on monthly means

For each year, a cubic spline is fitted to the monthly means of PC1 scores. Each functional observation corresponds to a year.



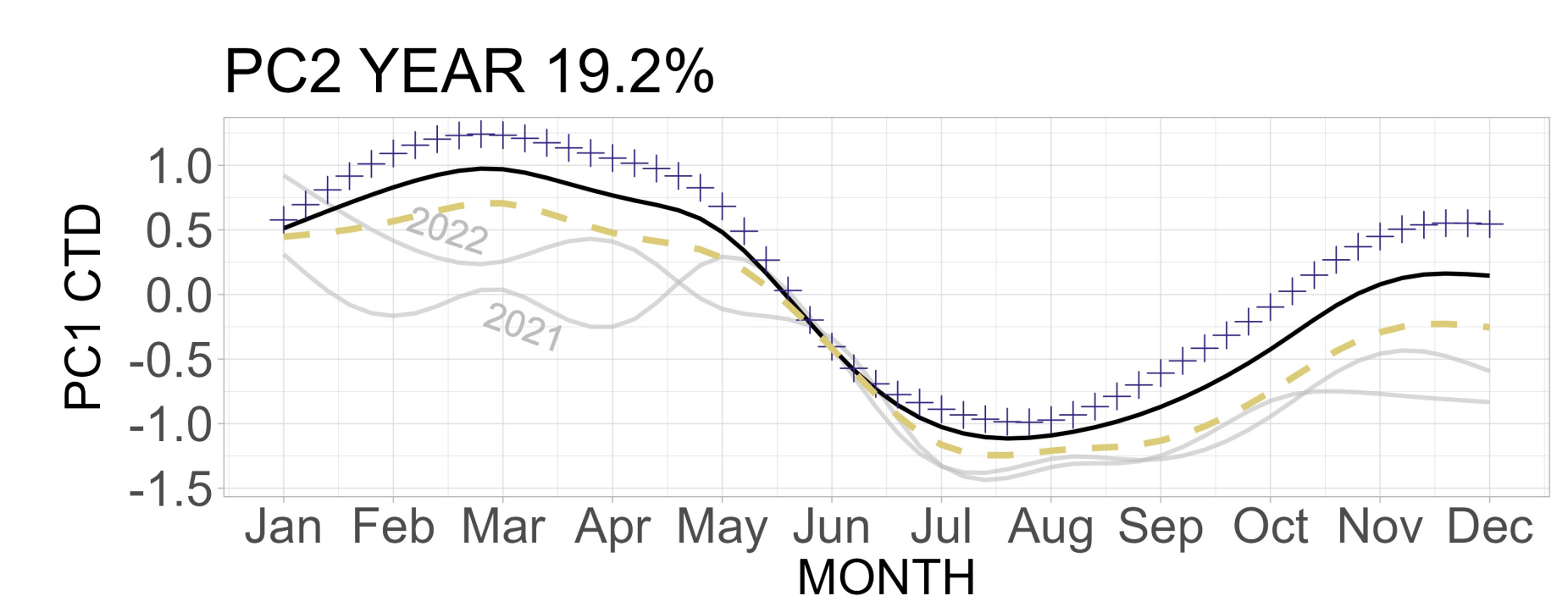
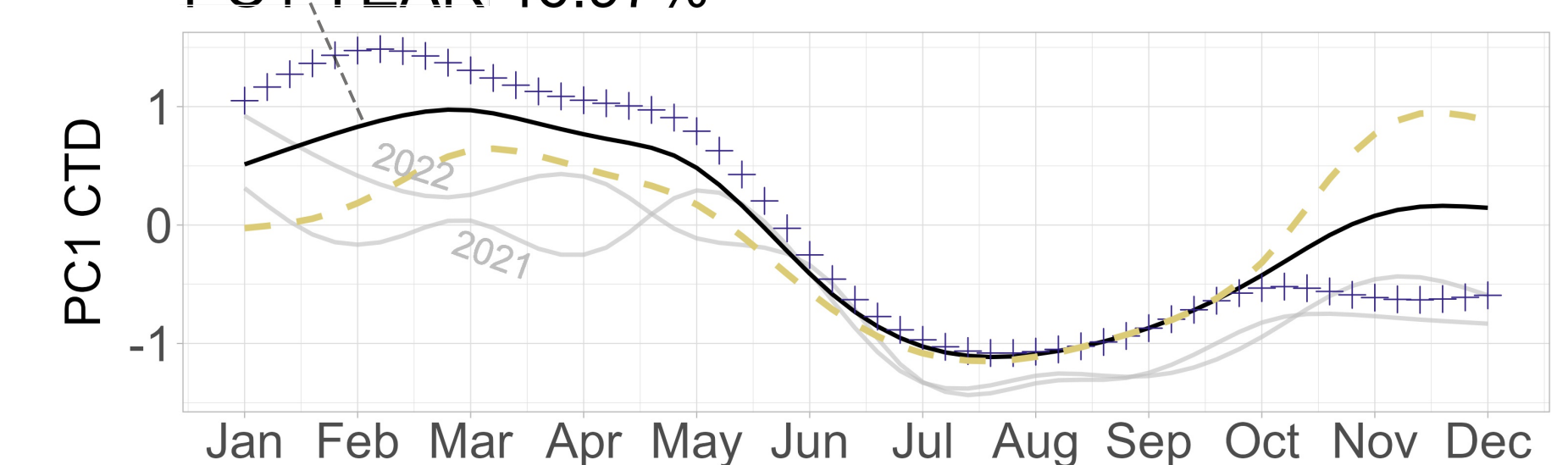
5 FPCA on year curves

Dimensional reduction of the smoothed monthly means throughout the year.



Fluorescence across the water column is higher than average from January to May, peaking in March. It then declines and remains below average until mid-October.

In 2021 and 2022, the fluorescence of the entire water column remained below average for the entire year, especially during the spring and autumn blooms.



Discussion

FPCA reveals that:

- Spring and autumn blooms are typically observed in Banyuls.
- In 2021 and 2022, the spring and autumn blooms were significantly below the average.

Meteo France reports that insolation at the Perpignan station—a city near Banyuls—was well below average in February 2021 and March 2022, with the latter setting a record for the lowest insolation duration for the month of March. The unusually low insolation during these months in 2021 and 2022 may have limited phytoplankton growth.

FDA to

- Take into account the continuous dimension of the CTD profiles
- Describe the evolution of the profiles over the year
- Identify years that show strong anomaly from the general pattern

References

- Karl, D., Church, M. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat Rev Microbiol* 12, 699–713 (2014). <https://doi.org/10.1038/nrmicro3333>
- Nerini, David & Manté, Claude & Monestiez, Pascal. (2022). Extending Functional Kriging When Data Are Multivariate Curves: Some Technical Considerations and Operational Solutions. 10.1002/9781119387916.ch4.