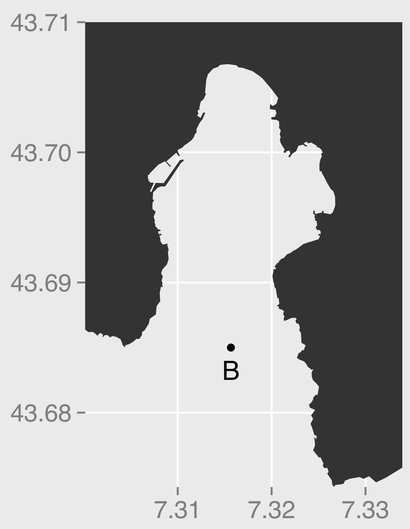
DETECTING HYDROLOGIC VARIATIONS IN A LONG TERM MONITORING TIME SERIES

Jean-Olivier Irisson, Alice Webb, Ornella Passafiume, Laure Mousseau
UPMC Univ Paris 06, UMR 7093, LOV, Observatoire océanologique, F-06234, Villefranche/mer, France
CNRS, UMR 7093, LOV, Observatoire océanologique, F-06234, Villefranche/mer, France
irisson@normalesup.org

INTRODUCTION



Long term time series are the most promising source of information to detect climate-induced changes and understand the modification of ecosystem processes.

Processes at the interface between hydrology and biology in the pelagic ecosystem are of particular interest in the time series collected at point B, in Villefranche's bay. Temperature and salinity are monitored weekly since 1957. More variables are recorded since 1995, using a CTD and water samples. Zooplankton is sampled daily since 1966 using nets.

Sampling this bay allows to capture both continental influences due to the proximity of the shore and pelagic processes because of the absence of continental shelf in this region of the Mediterranean. Furthermore, the Mediterranean itself is particularly sensitive to climate change.

DATA PREPARATION

Focus is put here on the richest data record, collected since 1995, at six depths, from surface to bottom. Of course, the record is not complete: missing values are present either sporadically or over long periods (years 2007 and 2008 for nutrients – Fig. 2).

We take advantage of the correlations between variables (chlorophyll and fluorescence, NO_3 and NO_2, temperature at 10 and 20 m depth, etc.) to reconstruct missing values in one record from existing values in others. Correlations between variables are captured through a Principal Component Analysis (Fig. 1).

The first 11 components of the PCA (80% of variance) are used to recompute the missing values. The PCA is then repeated on this new, complete, record until the predicted values converge. This technique allows to recreate credible seasonal cycles over two years of missing nutrients data, using depths 0 and 50 m where the data record is almost complete as well as correlations with salinity or chlorophyll.

Data is then regularised with a 7 days time step and the few completely missing weeks are linearly interpolated. This results in 883 data points per series.

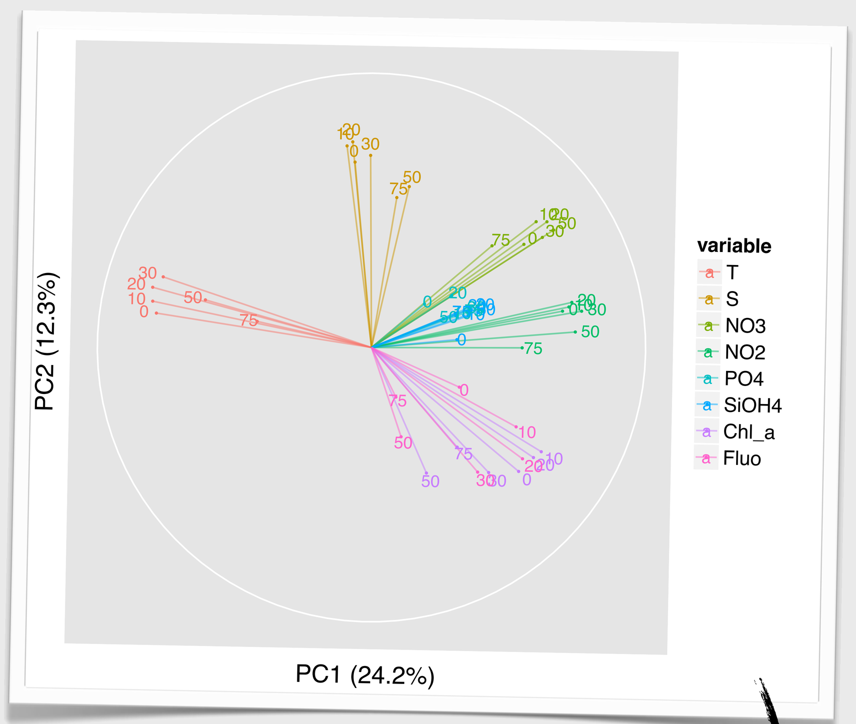


Figure 1: PCA factor space showing the correlations between hydrologic variables at six depths used to fill missing values.

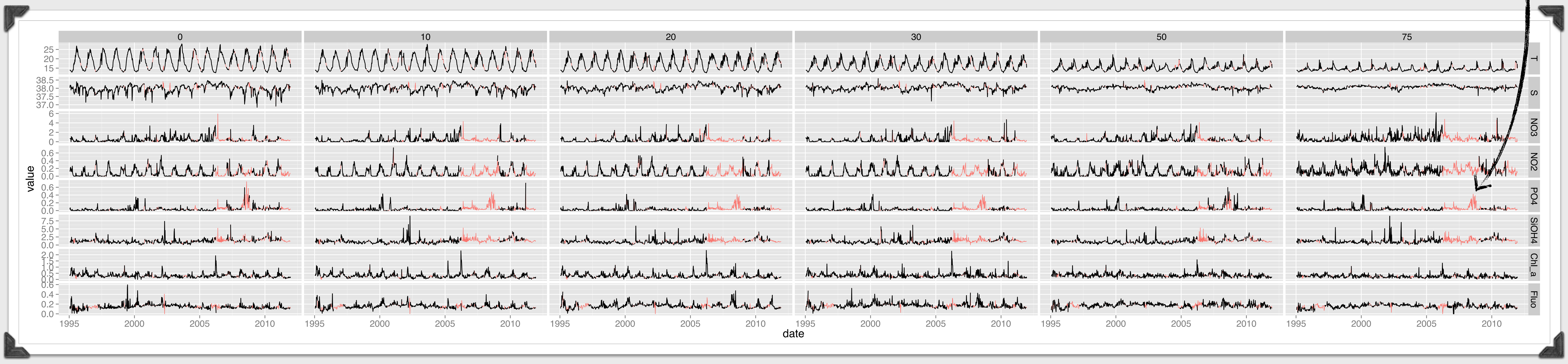


Figure 2: Complete data record in raw form (black) and after missing data imputation and regularisation (red). When data is present, the time step is already very regular, hence the almost perfect match even after regularisation.

DECOMPOSITION OF VARIANCE

The filled and regularised data series is influenced by a collection of processes acting on different scales. To tear them apart, we first use Eigen Vector Filtering (EVF). It decomposes each series into components representing orthogonal portions of the variance and sorted in decreasing order of importance (through a modified PCA procedure). The first EVF component captures seasonality in most series (Fig. 3), combined with larger-scale trends for salinity. Such a strong seasonality is expected in temperate systems and this result is therefore reassuring.

Because seasonality is the first source of variance in most series, it needs to be removed to see other processes. We locally fit two polynomials over short (3 months) and long (5 years) scales to extract the seasonal component and the larger scale trend (Fig. 4), through a process called STL (Seasonal-Trend decomposition based on Loess). Residuals sometime still have a seasonal component (Fig. 4) but removing it require to reduce the short scale to less than three months (i.e., a season) which introduces too much noise in the seasonal signal.

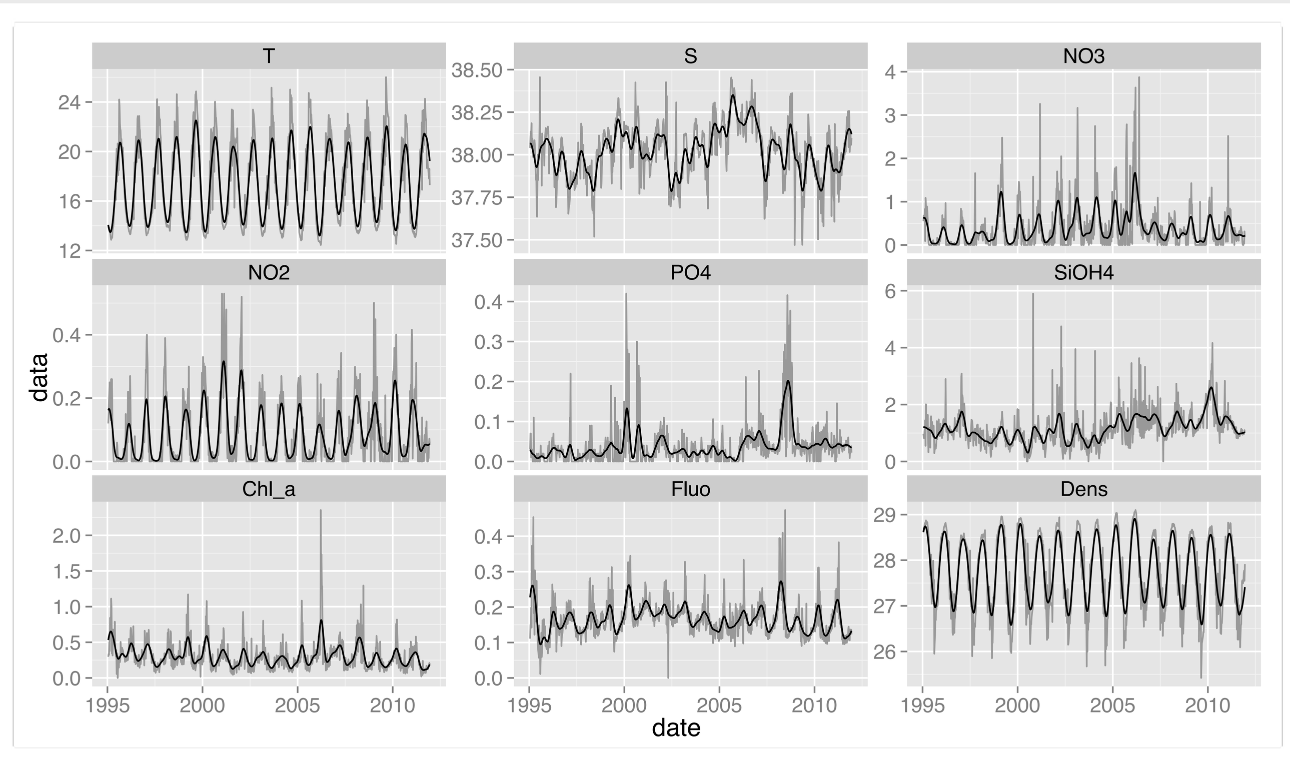


Figure 3: Data series at 20 m (grey) and first EVF component (black), which captures the process of highest variance (often seasonality here).

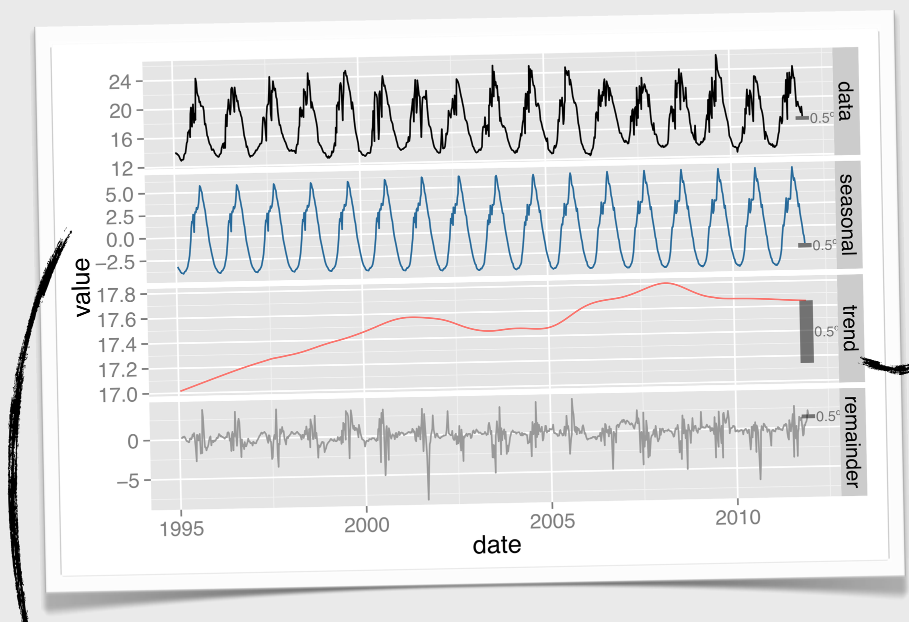


Figure 4: Seasonal-trend decomposition of temperature at 20 m.

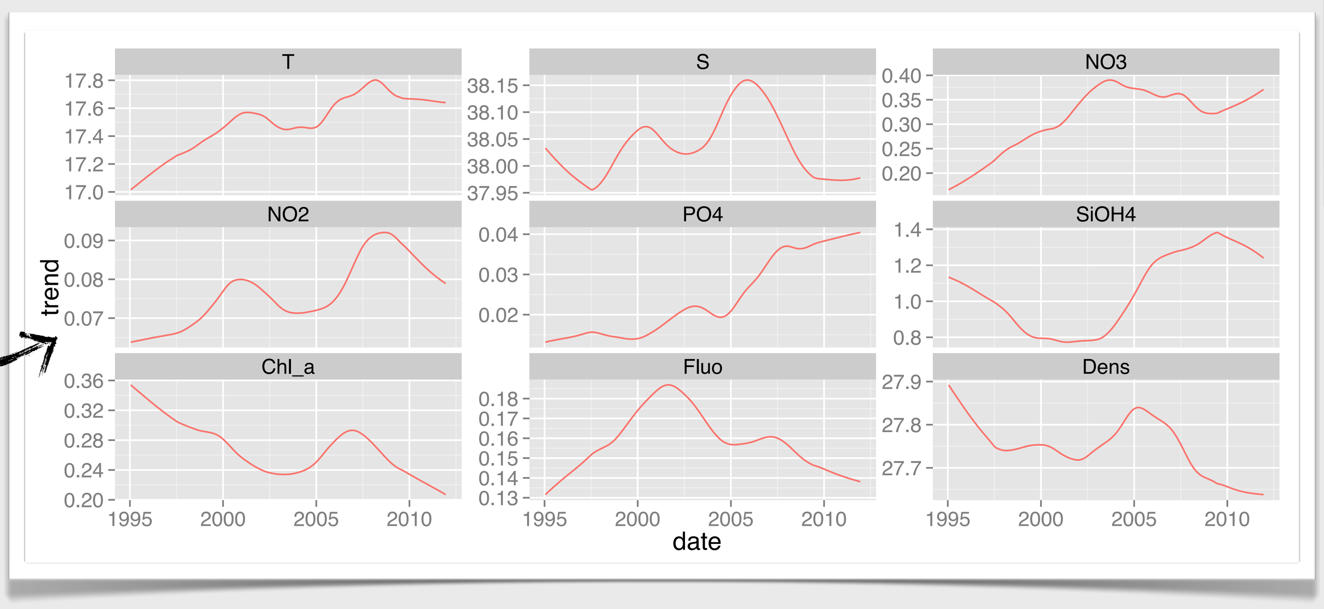


Figure 5: Trend component of the STL decomposition of data series at 20 m.

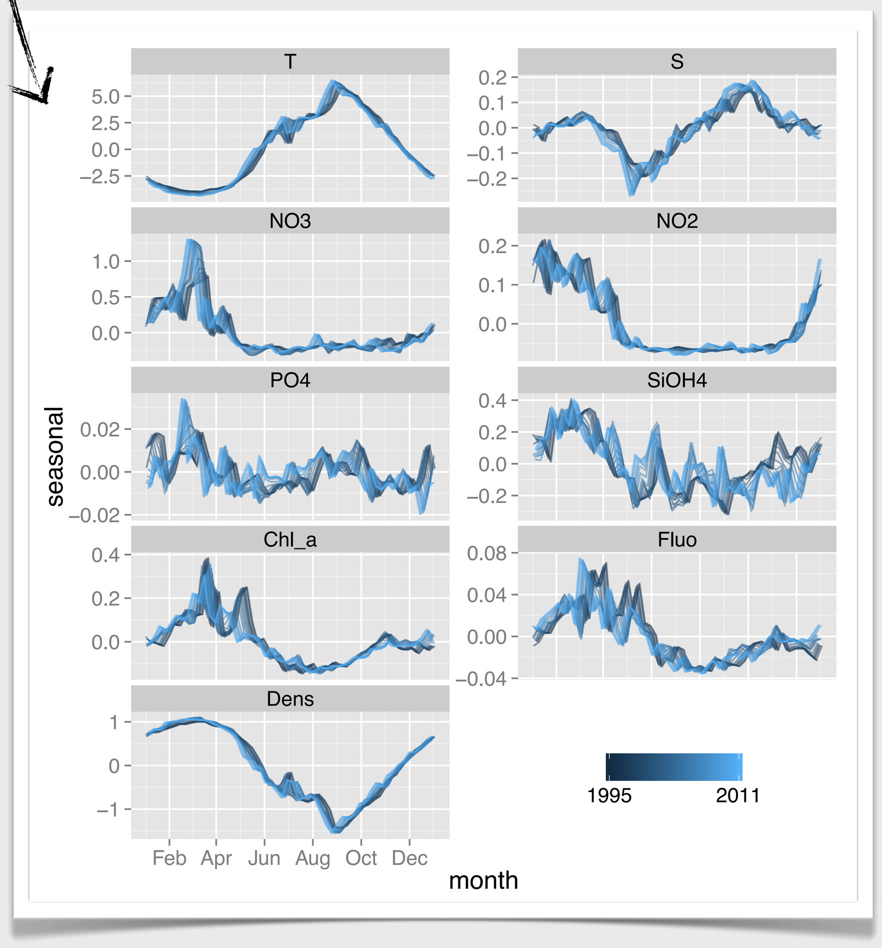


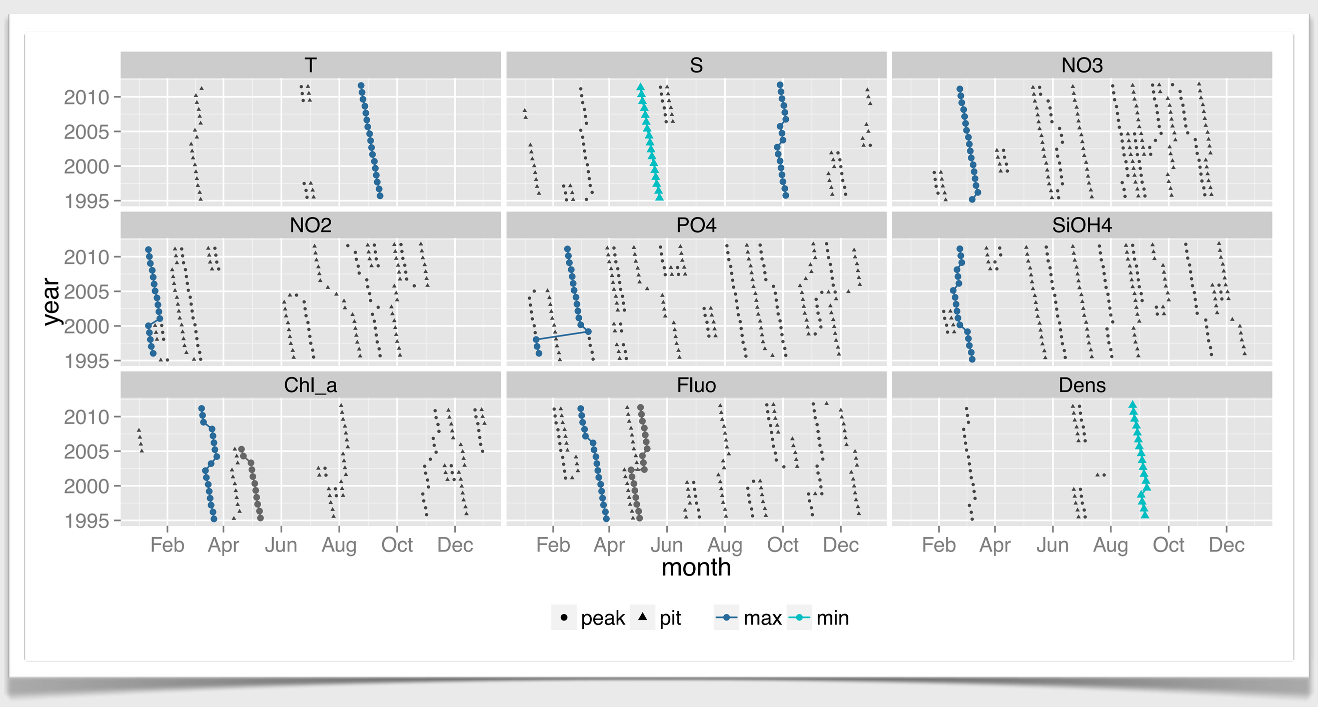
Figure 6: Seasonal component of the STL decomposition of data series at 20 m.

SEASONALITY

Looking at the superposition of seasonal components over a few years (Fig. 6), it seems that events are occurring earlier in the year. This is supported by the trajectory of remarkable peaks and pits through time which reveals the same shift towards earlier dates (Fig. 7). This trend is hardly visible in the raw data because of small scale inter-annual variability. While the steadiness of the shift might be exaggerated by the method, which smoothes out small-scale variability, the shift is quite clear.

The chlorophyll series also reveals a regime change, from two blooms, the main one in early spring and a weaker one in May, to just one in early spring (Fig. 7).

Figure 7: Dates of the peaks and pits in the seasonal components at 20 m. Maxima and minima are highlighted when relevant. The two main blooming events are shown for chlorophyll and fluorescence.



TRENDS

The trend for water temperature at 20 m shows an increase of 0.8°C over 17 years (Fig. 5). This fits with a larger scale trend in the whole data record (since 1957, not shown), although the increase accelerates. 0.8°C is already well above the global IPCC estimate of 0.74°C over the next 100 years. This increase is associated with a corresponding decrease in water density.

In the meantime, chlorophyll concentration, a proxy for primary production, decreases overall, while the concentrations of nutrients increases. This could suggest a top-down control of primary production by grazers rather than a bottom-up control by nutrients.

Yet, the chlorophyll trend peaked around 2007. This peak is associated with high salinity in the previous years, particularly 2006 (Fig. 5 and 2). During those years, dry and cold winters drove the increase in salinity. As a result, density also increased, winter convection was stronger and provided a more intense pulse of nutrients at the beginning of spring, hence feeding primary production. This pulse can be reliably be seen in 2006 (Fig. 2) but unfortunately not in the following years in which the nutrient series is mostly reconstructed.

CONCLUSIONS

- Iterative PCA can help fill large gaps in time series
- As expected, series are dominated by the seasonal component
- Temperature increased strongly in surface waters
- Primary production decreased, overall
- Dry years may cause an increase in salinity, mixing and primary production in the following year
- Remarkable events, such as blooms, seem to have shifted towards earlier dates in the year

Data provided by "Service Observation en MIlieu LITtoral-SORade, INSU-CNRS-UPMC, OOV". All analyses made in R 2.15.1, with packages stats and pastecs for time series decomposition, plyr and reshape2 for data management and ggplot2 for graphics.