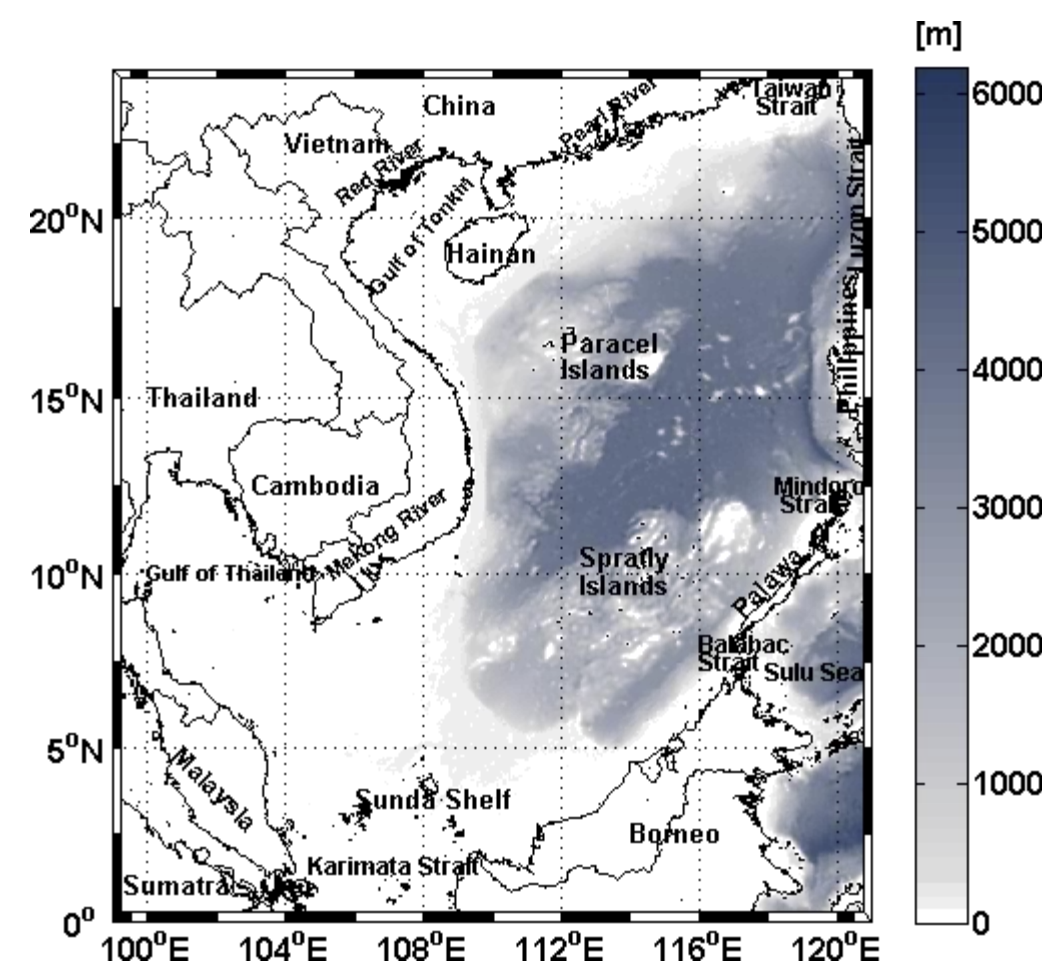


**Introduction** Chlorophyll-a (Chl-a) concentration can be considered as a good indication of phytoplankton biomass due to its universal presence in all species of phytoplankton, and one of the best available approaches to evaluate physical and biological interactions in the ocean. Insights of the Chl-a variability can help us better understand the dynamical phenomena in the seas and oceans such as upwellings, fronts, and eddies, etc. The aim of this study is to investigate the spatial and temporal variability of Chl-a in the South China Sea (SCS).

## Study Area

- SCS (**Fig. 1**) is connected to the East China Sea, the Sulu Sea, the Java Sea, the Indian Ocean, and the Pacific Ocean.
- The major rivers flow into the SCS: the Pearl, Red, and Mekong Rivers.
- SCS has complex bottom topography: wide and shallow continental shelves (<200 m) in the northwest and southwest, steep slopes without shelves in the east, and a deep basin (>5000 m) in the northeast.
- SCS is strongly influenced by the East Asian monsoon: cold northeast winds from November to April (~9 m/s); warm southwest winds (~6 m/s) from mid-May to mid-September.
- The surface circulations change under the influence of monsoons.

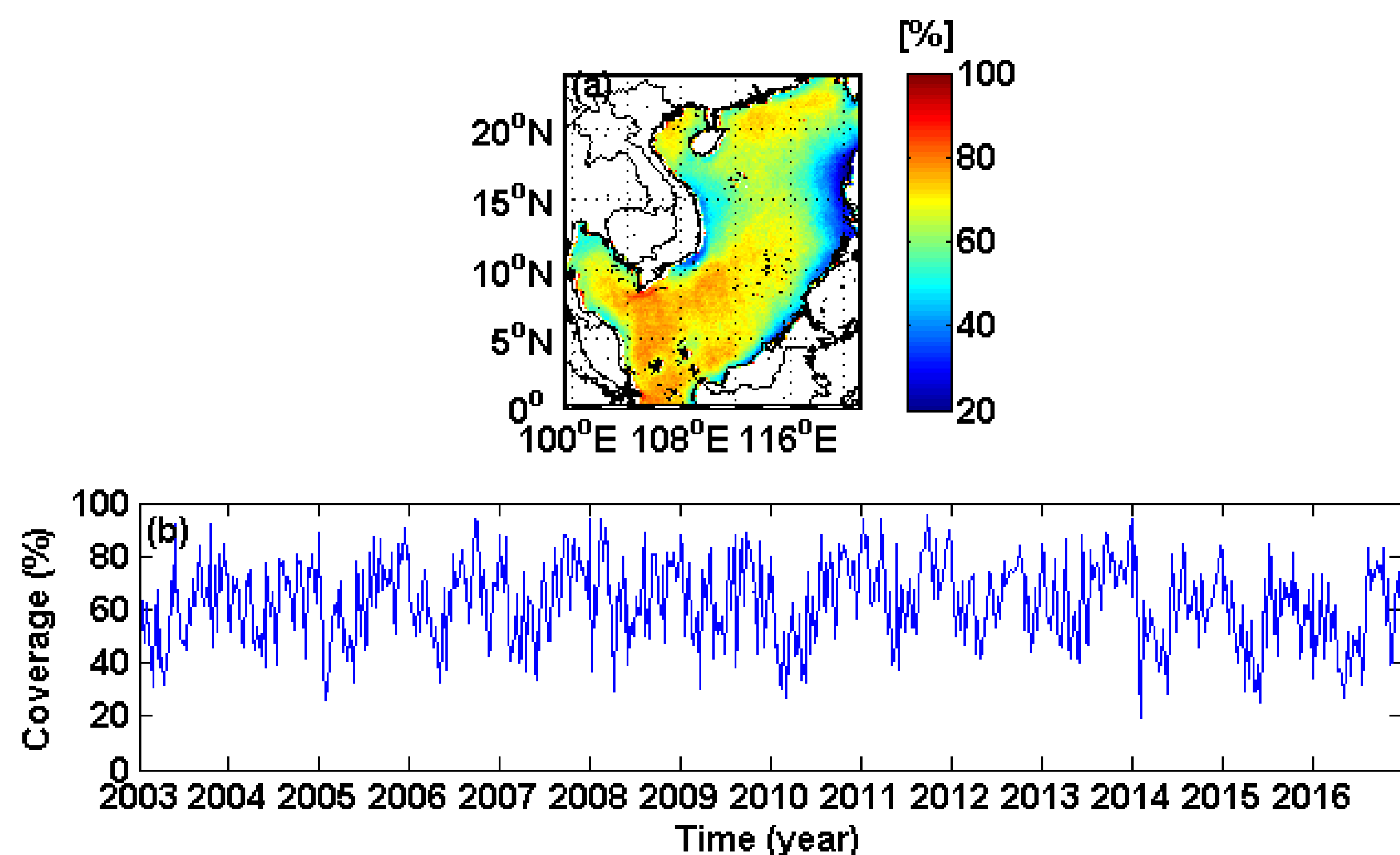


**FIGURE 1.** The South China Sea and its bathymetry (in m).

## Data and Method

**Data** 14-year (2003-2016) MODISA Chl-a was utilized in this study. The dataset consists of 644 gridded level-3, 8-day composite images with a resolution of 4 km (577 x 529 pixels). The MODIS sensor is, however, only capable of viewing in the visible and infrared wave bands; thus almost all images have gaps due to cloud cover (**Fig. 2**).

**Processing** Only images containing at least 5% of data were retained; pixels missing more than 95% of the time were considered as “land”.



**FIGURE 2.** Percentage of cloud-covered 8-day MODISA Chl-a in the SCS: a) spatial variation of cloud-covered Chl-a; b) temporal variation of cloud-covered Chl-a

**Method** To reconstruct cloud-covered satellite images, we used DINEOF (Data Interpolating Empirical Orthogonal Functions), described in Beckers et al. (2003), Alvera-Azcárate et al. (2005, 2009). Note also that Chl-a has non-Gaussian distribution; we used the base-10 logarithm transform before applying DINEOF on this data set (Campbell, 1995).

- The initial data input  $\mathbf{X}$  is obtained by subtracting the temporal and spatial mean, and setting the missing data to zero.
- A singular value decomposition of  $\mathbf{X}$  is performed to calculate the missing data:

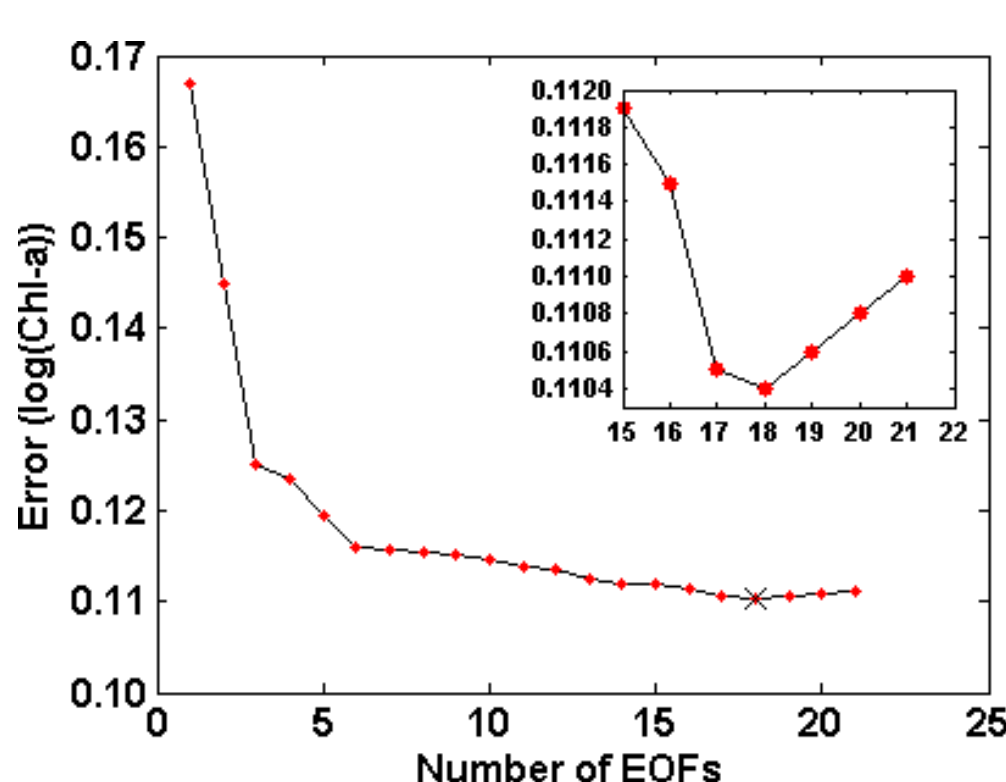
$$\mathbf{X}_{ij} = \sum_{p=1}^k \rho_p (\mathbf{u}_p)_i (\mathbf{v}_p)_j$$

where  $i, j$  are spatial and temporal indices, respectively;  $k$  is the retained number of EOF modes;  $\mathbf{u}$  and  $\mathbf{v}$  are the spatial and temporal modes, respectively; and  $\rho$  are the singular values. This step is repeated until a convergence is reached.

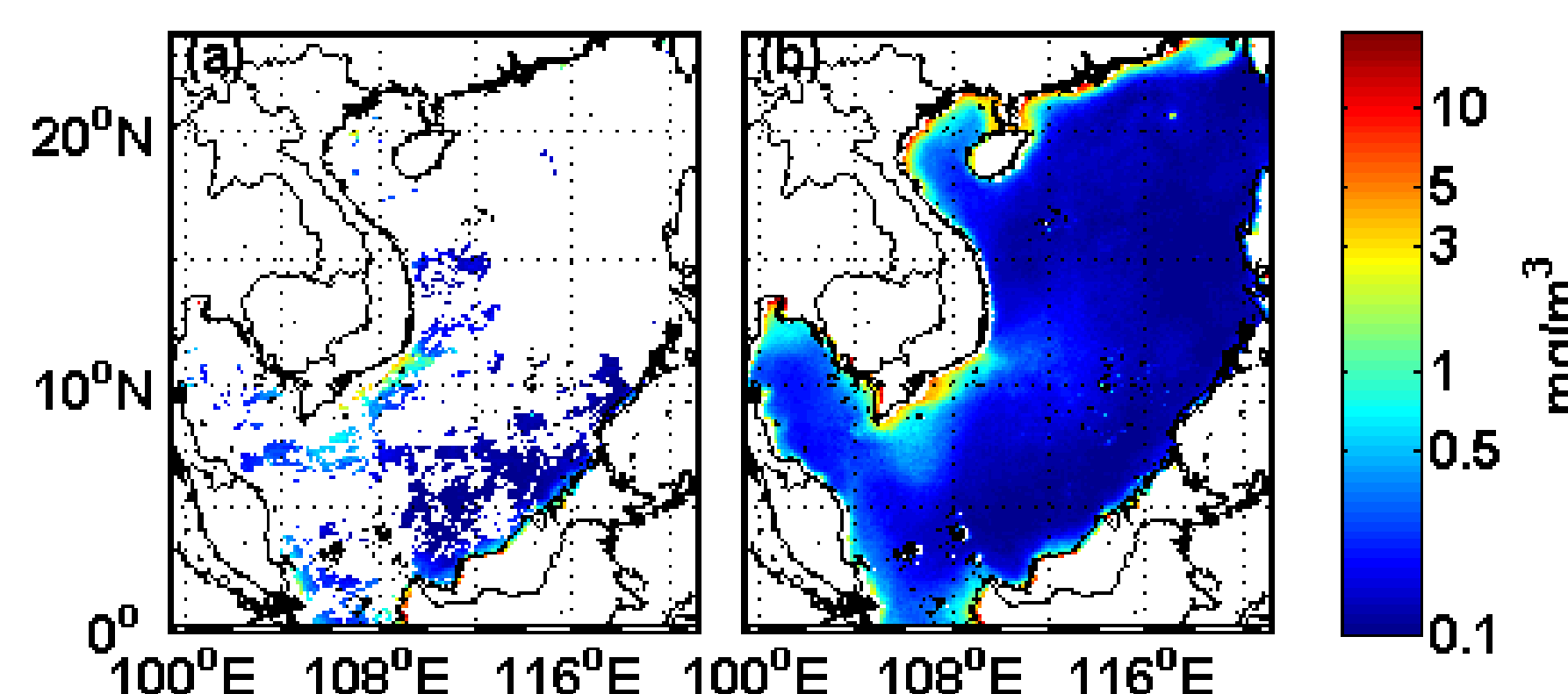
3) The optimal number of EOFs will be obtained when the global error between the reconstructed and cross-validation data points (3% of the dataset) is minimum (**Fig. 3**).

4) The optimal number of EOFs is used to reconstruct the whole matrix  $\mathbf{X}$

A sample of the reconstructed 8-day MODISA Chl-a is shown in **Fig. 4**



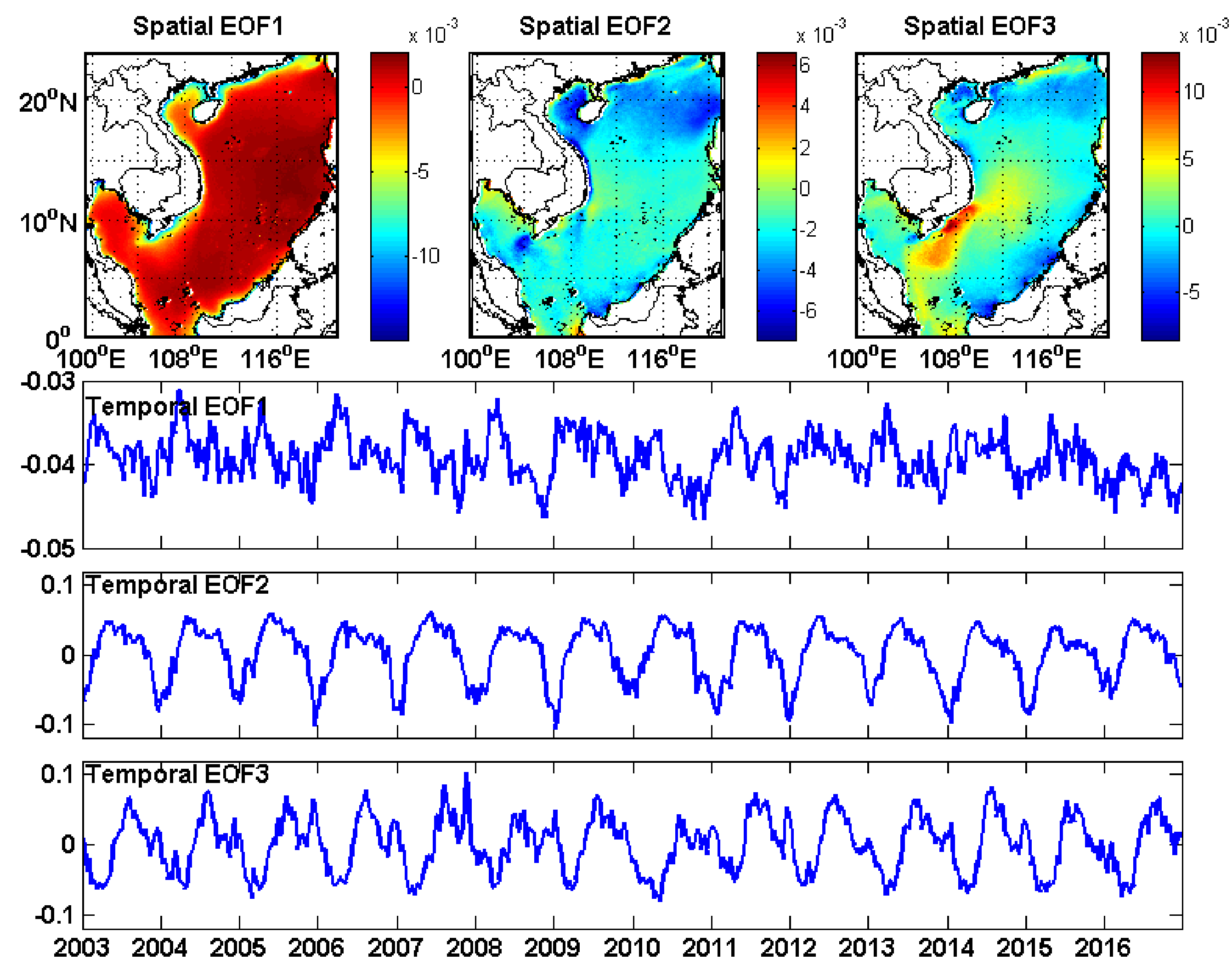
**FIGURE 3.** Expected errors obtained from cross-validation. The embedded figure on the upper-right corner zooms in the error with the optimal number of EOFs



**FIGURE 4.** An example of the reconstructed 8-day MODISA Chl-a from 12 to 20 August 2016: a) the original MODISA Chl-a, b) the reconstructed Chl-a.

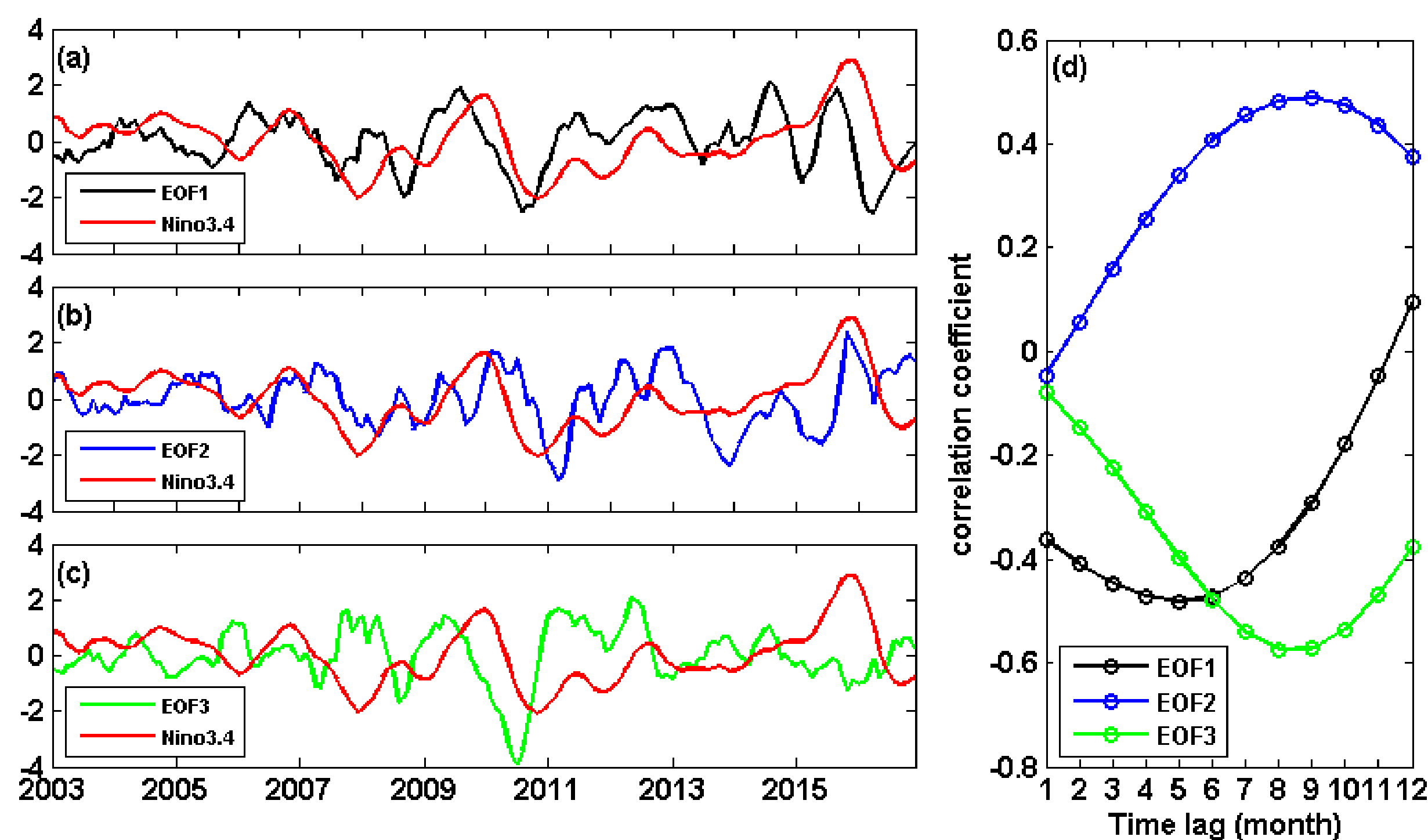
## EOF Analysis

We analyze the three most dominant modes with 78.05%, 9.78%, and 2.76% of the total variance in **Fig. 5**.



**FIGURE 5.** The first three modes of the reconstructed Chl-a. Note that Chl-a is on a base10 logarithm scale

## Correlation with ENSO



**FIGURE 6.** Correlation between the monthly anomalies of the first three temporal EOFs and Nino3.4 SST region (5°N-5°S, 120°-170°W) (a, b, and c). For each time-series, data are smoothed by 5-month running mean and the linear trend over 14 years is removed. Note also that each time-series is normalized. d shows the correlation coefficients between the first three temporal EOFs and Nino3.4 index versus time lags. All correlation coefficients are significant at the 95% confidence level.

## Conclusions

- EOF1 presents the high Chl-a concentrations in the coastal regions, except those of the Palawan and Philippines. Generally, this pattern has peaks in summer (June-July) and winter (November-December).
- EOF2 shows the seasonal variability of Chl-a in the whole basin. The Chl-a concentrations increase in winter and decrease in summer, with the highest variability in the west of Luzon Strait, the east of Tonkin Gulf, the tip of southern Vietnam, the Borneo coast, and a narrow region along the northeast of Vietnam coast.
- EOF3 highlights the out-of-phase variability of the southern SCS Chl-a between the west and east coasts in winter and summer, coinciding with the variability of the southern SCS SST (Huynh et al. 2016).
- The variability of Chl-a is under the influence of ENSO with a time lag of 5-9 months (**Fig. 6**).

### Acknowledgements

8-day MODISA Chl-a data were obtained from OceanColor Web (<https://oceancolor.gsfc.nasa.gov>). Nino 3.4 index was downloaded from <http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices>.

### Main references

- Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J.-M. (2005). Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions: application to the Adriatic sea surface temperature. *Ocean Modelling*, 9(4):325-346
- Alvera-Azcárate, A., Barth, A., Sirjacobs, D., and Beckers, J.-M. (2009). Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF. *Ocean Science Discussions*, 6(2):1547-1568.
- Beckers, J.-M. and Rixen, M. (2003). EOF calculations and data filling from incomplete oceanographic datasets. *J. Atmos. Oceanic Technol.*, 20(12):1839-1856.
- Campbell, J. W. (1995). The lognormal distribution as a model for bio-optical variability in the sea. *Journal of Geophysical Research: Oceans*, 100(C7):13237-13254.
- Huynh, H.-N. T., Alvera-Azcárate, A., Barth, A., and Beckers, J.-M. (2016). Reconstruction and analysis of long-term satellite-derived sea surface temperature for the South China Sea. *Journal of Oceanography*, 72(5):707-726.