Progress In Oceanography
May 2015, Volume 134 Pages 221-231
https://doi.org/10.1016/j.pocean.2015.02.001
https://archimer.ifremer.fr/doc/00668/78012/

Moving toward finer scales in oceanography: Predictive linear functional model of Chlorophyll a profile from light data

Bayle Severine [1], Monestiez Pascal [1], Guinet Christophe [2], Nerini David [3, *]

[1] INRA, Biostat & Proc Spatiaux BioSP, UR546, F-84914 Avignon, France.
[2] CEBC CNRS, F-79360 Villiers En Bois, France.
[3] Aix Marseille Univ, Pytheas Inst OSU, UMR 7294, Mediterranean Inst Oceanog, F-13288 Marseille 09, France.

* Corresponding author : David Nerini, email address : david.nerini@univ-amu.fr

severine.bayle@avignon.inra.fr ; pascal.monestiez@a-vignon.inra.fr ; christophe.guinet@cebc.cnrs.fr

Abstract :

The Southern Ocean plays a key role in ocean atmosphere carbon dioxide fluxes. Estimation of carbon exchanges between ocean and atmosphere must rely on accurate estimations of primary productivity which require measurements of phytoplankton concentration within the water column. In this paper, we are interested in relationships between primary productivity and light in the Antarctic ocean. The originality of this work is twofold. Starting from physical hypothesis, a statistical model is constructed for the prediction of Chlorophyll a (Chl a) profiles where light profiles are used as a covariate. Taking into account of the functional nature of the data, solutions are proposed to estimate continuous vertical profiles from discrete data sampled by elephant seals equipped with a new generation of oceanographic tags. Bootstrapped prediction intervals show a good quality of prediction of Chl a profiles, giving access to the shape of the profiles along depth and to the submesoscale structure of phytoplankton within the euphotic layer of the Southern Ocean.

Highlights

► Prediction of Chlorophyll-*a* profiles is achieved in Antarctic ocean. ► A functional linear model is constructed using light curves as covariate. ► The predictive capabilities of the model associated to confidence intervals show that it is possible to predict Chlorophyll-*a* at fine scale. ► Chlorophyll-*a* prediction at fine scale highlights sub-mesoscale variations.

1. Introduction

Marine phytoplankton contributes to roughly half of the biosphere's primary production and therefore represents a vital link between living and inorganic stocks of carbon (Behrenfeld et al., 2006). But there is conflicting evidence on how this biological productivity will respond to global warming and climate change, particularly in the Southern Ocean, which plays an essential role in the carbon cycle. In that context, the measurement of Chl *a* concentration in the waters of the Southern Ocean is a crucial indicator of the amount of phytoplankton and of the spatial and temporal variability of primary productivity (Behrenfeld and Falkowski, 1997) and must enable a better quantification of CO2 fluxes.

However, the understanding of both the primary production variability and its spatial structure at submesoscale is hampered by the lack of *in situ* observations. Furthermore, the degree of confidence for observations of primary production derived from satellite-based estimates of phytoplankton biomass is still open to debate, especially with regard to the Southern Ocean (Guinet et al., 2013b). There is evidence of major limitations regarding the use of satellite assessment of primary production within the Southern Ocean. Satellites scan the sea surface, while deep fluorescence maxima can be found at depths of 40 and 75m within the frontal zone of the Antarctic Circumpolar Current (Queguiner and Brzezinski, 2002). Primary production cannot be properly assessed due to persistent cloud cover that precludes satellite detection of ephemeral phytoplankton blooms (Arrigo et al., 1997; Buesseler and Boyd, 2003). A better description of spatial (horizontal and vertical) and temporal (seasonal, inter-annual) distribution of phytoplankton is therefore essential to understand how primary production within the Southern Ocean and therefore CO2 fluxes will respond to climate change.

Phytoplankton concentration is generally quantified through active measurement of Chl *a* fluorescence. However few autonomous platforms such as ARGO floats or gliders are used to collect this data in the ocean especially (Fedak, 2013).

In recent years, a number of broad ranging deep diving marine predators have been equipped with electronic tags to investigate their foraging ecology and sample *in situ* oceanographic variables over broad sectors of the ocean (Charrassin et al., 2008; Boehlert et al., 2001; Fedak et al., 2002; Block et al., 2002; McMahon et al., 2005; Biuw et al., 2007; Charrassin et al., 2008). Oceanographic variables sampled with animal-borne electronic tags include

temperature and salinity, and more recently fluorometers were integrated in order to obtain along with temperature and salinity simultaneous estimates of Chl a in the water column as experienced by pelagic animals (Xing et al., 2012; Guinet et al., 2013b).

However due to high energy requirement only few profiles can be sampled daily. Recently, using Southern elephant seals simultaneously equipped with a fluorometer and a light logger, Jaud et al. (2012) showed that light attenuation was strongly correlated with Chl a concentration measured by the fluorometer within the euphotic layer. In a pioneering work, Teo et al. (2009) using the bio-optical model of Morel (1988) demonstrated that in situ Chl a concentration profiles estimated from fluorescence measurements could be estimated from monitoring light level and depth data collected by electronic tags.

The proportionality of Chl a and fluorescence is known to be modulated by the taxonomic composition and physiological acclimation mechanisms essentially related to light. Among the physiological acclimation mechanisms affecting the fluorescence-Chl a relationship, there is the depression of the fluorescence signal in surface waters during daylight and especially at maximum solar elevation. The so-called fluorescence quenching is the most obvious phenomena (Marra, 1997; Holm-Hansen et al., 2000; Xing et al., 2012). Fluorescence quenching does indeed represent a collection of different photoprotective mechanisms to avoid photodamage under excessive sunlight energy (Kiefer, 1973; Maxwell and Johnson, 2000).

To our understanding, the fluorescence profiles collected in Teo et al. (2009) were not quenching-corrected and this could have resulted in an underestimation of Chl a concentration within the first 30 to 60 m of the water column from the fluorescence measurements. This, plus other factors such as the presence of non-phytoplanktonic particles such as zooplankton which might contribute to light attenuation, could explain the poor performance of the prediction of the Chl a profiles from light data under certain conditions, and Teo et al. (2009) recognized the need to improve their method.

Following up these two studies, we propose to construct a functional linear model in order to predict in situ Chl a concentration profiles using light profiles as a predictive variable. Compared to the previous studies, the originality of this work is to include into the model the functional nature of the data. Indeed, the dataset used to construct the model has been sampled by elephant seals equipped with tags. In the course of an elephant seal trajectory, each data profile arrives as a discrete set of observations of light and

fluorimetry sampled at varying depths. The first section of this paper is devoted to problems of data sampling with marine mammals for light and Chl *a* variables. As for the bio-optical model described in Teo et al. (2009), the construction of the linear functional model starts from the Beer-Lambert relationship defined by Bouguer (1729) which links the light absorption to properties of an homogeneous environment. This is the objective of the second section : bring the construction of the statistical model by linking Chl *a* profiles to light profiles under more realistic assumptions than those of Beer-Lambert. We will focus more particularly on the fact that the model is constructed using the derivative of log light profiles. This section also includes technical solutions for parameters estimation, construction of observed profiles from elephant seals dataset, including constraints over the shape of profiles, construction of bootstrap prediction intervals and measures of model accuracy. The third section present the main results. The paper is endded with a brief discussion.

2. Elephant seal Dataset

In October 2009, at the beginning of the austral spring, 3 post-breeding Southern elephant seal females from Kerguelen Island were anesthetized by intravenous injection of tiletamine and zolazepam 1:1 to be fitted with a Satellite Relayed Data Logger sampling pressure, temperature, salinity and fluorescence (CTD-Fluo SRDL afterward) -developed by the Sea Mammal Research Unit, St Andrews University, Scotland, in collaboration with the Centre d'Etudes Biologiques de Chizé- combined with an MK9 (Wildlife Computer, USA) time depth recorder (TDR) glued on the back of the CTD-Fluo SRDL next to the fluorometer with both the fluorometer and light sensor facing backward. The location of the light sensor at the back or at the front of the satellite does not change the relationship between light and fluorescence as mentioned in Jaud et al. (2012). The package was then glued on the fur of the Southern elephant seal's head using a two component industrial epoxy (Araldite AW 2101).

Through the whole post breeding foraging trip, fluorescence was generally measured twice a day from the CTD-Fluo tags while light and pressure was monitored continuously at 2 hz by the logger. The CTD-Fluo SRDL included a Keller type pressure sensor (series PA7 0 to 2000 dbar $\pm$ 1 dbar), a fast response Platinum Resistance Thermometer (PRT) (-5°C to 35°C $\pm$ 0.005°C, 0.7 seconds response time), an induction conductivity sensor devel-

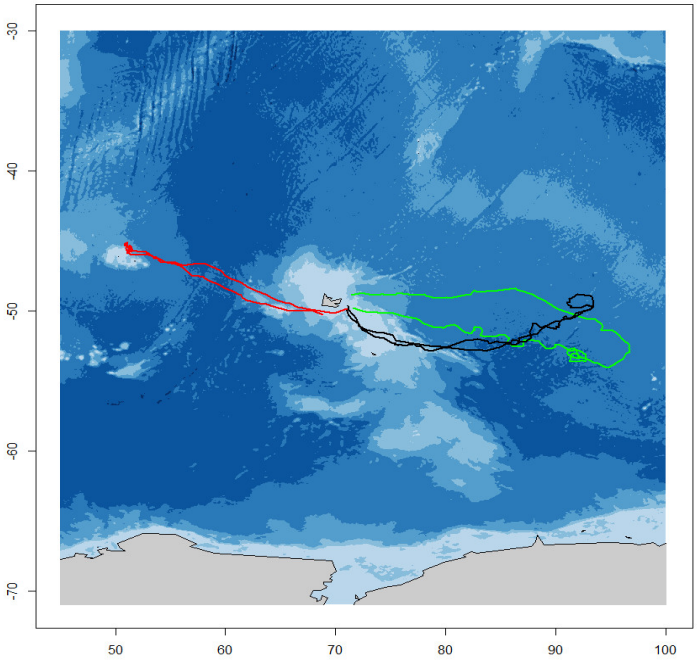


Figure A: Paths of the three elephant seals. Starting from and turn back to Kerguelen Islands (center), animals turn back to land after a journey of 3 months, with Antarctic continent at the bottom (green: path 1, black: path 2, red: path 3).

oped by Valport (UK, range: 0 to 80 mS cm^{-1}, accuracy: better than 0.02 mS cm^{-1}), and a Cyclops 7 fluorometer from Turner Design with a dynamic range set between 0 to 5μg of Chl *a* (chl-a.L^{-1}). The MK9 TDR loggers were set to sample depth (0 to 1500 $\pm$ 1 meter), water temperature (-40°C to +60°C $\pm$ 0.1°C) and light (5.10^{-2} W.cm^{-2} to 5.10^{-12} W.cm^{-2} in blue Wavelength) every two seconds. The MK9 integrated light sensor was tested under laboratory condition and was found to exhibit its highest sensitivity at 465 nm, with a mean sensitivity range of $405 - 480$ nm (Vacquié-Garcia, 2014). The wavelength correspond to blue light. Blue light is the least water attenuated wavelength, and therefore the blue light is able to reach greater depth encompassing the whole euphotic layers. A complete description of these tags is available in Boehme et al. (2009). Light values are converted

inboard via a log treatment to reduce the light measurements to a 3 digit value.

Before deployment on Southern elephant seals, each CTD-Fluo SRDL was calibrated at sea during the BOUSSOLE campaign (Guinet et al., 2013b) by comparison with in situ measurement from Niskin bottles. A coefficient was calculated for each tag to convert the fluorescence values to an actual Chl *a* concentration.

When CTD-Fluo SRDL were deployed on elephant seals, Chl *a* concentration (μg.l^{-1}) derivated from fluorescence measurements was assessed continuously at a two second sampling rate for the last 180 meters of the ascent phase of the dive. The depth of 180m was selected as the threshold because it encompasses the euphotic layer, which is generally close to 150m. Each profile, transmitted via the ARGOS system, consists of a maximum of eighteen sections of ten meters. The average fluorescence value is associated with the median depth of each segment (-5 to -175m). About two fluorescence profiles are sampled and transmitted daily by these tags via the ARGOS system. Temperature and salinity were treated similarly for the first 180m, and six measurements were made at depths exceeding 180 to ensure that the best reconstruction of the high resolution temperature and salinity profile might also be transmitted.

Daylight fluorescence profiles are affected by quenching, defined as photoinhibition due to an excess of light, resulting in an artificial deep maximum Chl *a* concentration. In well mixed waters representing about 84% of available profiles, the fluorescence profiles obtained during daylight hours were post-processed to correct the quenching effect according to Xing et al. (2012). Daylight profiles obtained in well stratified water (less than 20%) were excluded from the analysis as quenching could not be corrected accurately under these conditions (Xing et al., 2012). Processed Chl *a* data including the 3 individuals (tags 11259, 11260 and 11263) used in this study are freely downloadable at doi:10.7491/MEMO.1 (Guinet et al., 2013b).

The 3 tags were recovered in January 2010 at the beginning of the austral summer when Southern elephant seal females came back to Kerguelen Island to moult and MK9 pressure and light data were downloaded.

In this study, only complete Chl *a* profiles (i.e. 18 data points) were used for a better estimation and smoothing of Chl *a* profiles. So, among the 436 Chl *a* profiles sampled, 407 were included and among them we selected those sampled only during daylight hours with a sun angle greater than 20° degrees above the horizon to match them to the corresponding the light data

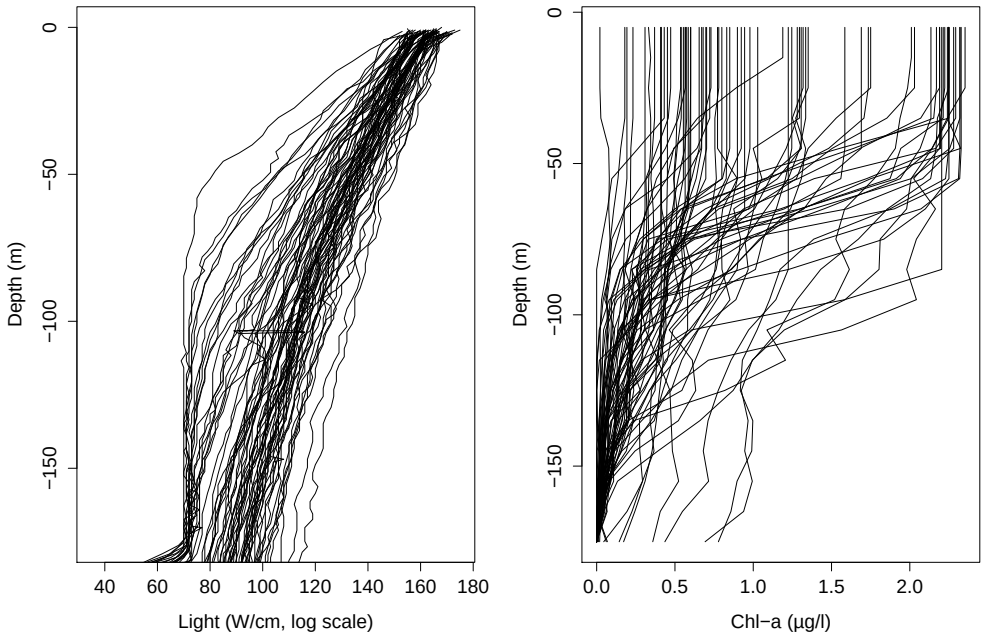


Figure B: Examples of light and Chl-a raw data sampled on the path of the elephant seal number 1. The objective is to construct matched observed profiles using these pointwise observations.

profiles sampled during the ascent phase. A complete data set is composed of $n = 208$ pairs of Chl a and light data profiles (Tab. A).

3. Statistical methods

3.1. Constructing the functional linear model

Let $C(z)$ be the Chl a variable and $L(z)$ the light variable. These variables are functions with argument z ranging from subsurface $Z_m = 5m$ to maximum depth $Z_M = 175m$ in the interval $[Z_m; Z_M]$. We wish to investigate to what extent Chl a profiles can be predicted from information contained in light profiles using a collection of pairwise functions $(L_i, C_i), i = 1, \ldots, n$. This sample of observed profiles must be constructed from pointwise data sampled by the elephant seals.

Usually, a light profile in a homogeneous and weakly concentrated liquid

ARGOS number	Number of elephant seal	Total number of registered Chl *a* profiles with 18 observations	Number of considered Chl *a* profiles
11260	1	142	73
11263	2	148	73
11259	3	117	62

Table A: Different elephant seals with number of available profiles. The last column is the number of Chl *a* profiles with 18 observations sampled during daylight hours.

medium is constructed using the Beer-lambert equation (Bouguer, 1729). The intensity of light is supposed to decrease exponentially in accordance with depth following the relationship

$$L(z, \lambda) = L_0(\lambda) * \exp(-q(\lambda) * z)$$

where L is the light intensity, z is depth, λ is a wavelength, L_0 represents the light intensity at the surface, and q is the light attenuation coefficient. Remind that in our case, only the blue wavelength has been sampled. Then, considering a fixed value of λ (blue wavelength in this study), this last equation is the solution of the following ordinary differential equation

$$\frac{dL}{dz}(z) = -qL(z),\ L(z = Z_m) = L_0.$$

The Beer-Lambert assumptions suggest that the profile of Chl *a* concentration is constant alongside depth. In first approximation, it can be considered that the Chl *a* concentration is independent from depth and proportional to the coefficient q such that

$$C(z) \propto q.$$

However, both for light and Chl *a*, the hypothesis are in contradiction with the observations of the water column (Fig. B). Light profiles do not decrease exponentially with depth and Chl *a* profiles are not linear with depth.
One way to release the hypothesis of homogeneous environment to broader assumptions is to consider that the coefficient of light attenuation is depending with depth. This implies that the Chl *a* concentration is proportional to $q(z)$ *i.e.*

$$C(z) \propto q(z)$$

and also that the light attenuation is solution of the following differential equation

$$\frac{dL}{dz}(z) = -q(z)L(z),\ L(z = Z_m) = L_0.$$

The solution of the last equation is given with

$$L(z) = L_0 \exp\left(-\int_{Z_m}^{z} q(s)ds\right),\ L(z = Z_m) = L_0, \tag{1}$$

which is a monotonic function : the light decreases as depth increases. Taking the derivative of that solution in logarithmic scale, we obtain

$$L'(z) = \frac{d\log(L(z))}{dz} = q(Z_m) - q(z),$$

the derivative of the log-light is then proportional to the attenuation coefficient $q(z)$. As $C(z) \propto q(z)$ and $L'(z) \propto q(z)$, one can properly assume that $C(z) \propto L'(z)$.

The most simple linear model that can be constructed to predict the Chl a functional variable with $L'(z)$ used as a covariate reads

$$C(z) = \alpha + \beta L'(z) + \varepsilon(z),$$

where α and β are real parameters that must be estimated with the sample, and $\varepsilon(z)$ is a remainder. A more general version of this simple model can be constructed by considering possible variations of parameters α and β alongside the depth :

$$C(z) = \alpha(z) + \beta(z)L'(z) + \varepsilon(z).$$

However, this model only considers interactions between variables at the same depth. From a predictive point-of-view, it can be relevant to construct a more flexible model which gives predictions of the Chl a at depth z by using the information of the whole curve L'. For that purpose, we consider the following linear functional model

$$C(z) = \alpha(z) + \int_{Z_m}^{Z_M} \beta(s, z)L'(s)ds + \varepsilon(z). \tag{2}$$

The bivariate coefficient $\beta(s,z)$ can be interpreted for a fixed value of z as a loading function which gives the ability to take into account cross-dependencies between Chl a and light at different depths. The coefficient $\alpha(z)$ is a function that acts as an intercept. The last term $\varepsilon(z)$ is considered as a random error term. The properties of such a model and different ways to estimate the parameters have already been studied in Ramsay and Silverman (2005) and related references therein. In this study, both coefficients $\alpha(z)$ and $\beta(s,z)$ are estimated using the data samples at hands (see Appendix A). The advantage of such a model is that it includes the more simple versions of the linear model presented above but allows more flexibility if required.

3.2. Constructing Chl a and light profiles

Consider a sampled profile of variable C which arrives as p discrete observations $(z_j, c_j), j = 1, \ldots, p$. We wish to reconstruct this unknown observed profile $C(z)$ using th pointwise observations (Fig. B). One way to proceed is to consider that an observed profile $C(z)$ is expressed as a linear combination of known basis functions $\phi_k, k = 1, \ldots, K$ such that

$$c_j = C(z_j) + \epsilon_j = \sum_{k=1}^{K} a_k \phi_k(z_j) + \epsilon_j.$$

The deterministic part $C(z) = \sum_{k=1}^{K} a_k \phi_k(z)$ of the data is entirely determined by the coefficients a_k which are estimated when minimizing the sum of squares of the errors ϵ_j. In our case, releasing all assuptions on the shape of the curve, we choose a B-spline basis, which is defined as a piecewise polynomials of order 4 where the coefficients are computed when minimizing the penalized cost function

$$1/p \sum_{j=1}^{p} \left(c_j - C(z_j)\right)^2 + \theta \int \left(C''(u)\right)^2 du.$$

The smoothing parameter θ gives the trade-off between smoothness of the curve (norm of its second derivative) and the data fitting (Fig. C). The smoothing parameter θ and the number K of basis functions can be chosen by cross-validation (Craven and Wahba, 1979; Hosseini-Nasab, 2012). Once the polynomial regression is achieved for both observed variables L and C, we dispose of a sample of matched functions $\{(L_i, C_i), i = 1, \ldots, n\}$ for Chl a profiles and for light profiles. Each of these functions are entirely

determined by the knowledge of their coefficients when expanded into the basis. The advantage of considering a smoothing spline basis expansion is that the derivative of any curve can be explicitly calculated, some required constraints (positivity, monotony, ...) can be included as well. We will show in the following section how to proceed in order to select the right number of basis functions and to impose monotony constraints for light profiles fitting as suggested by equation 1.

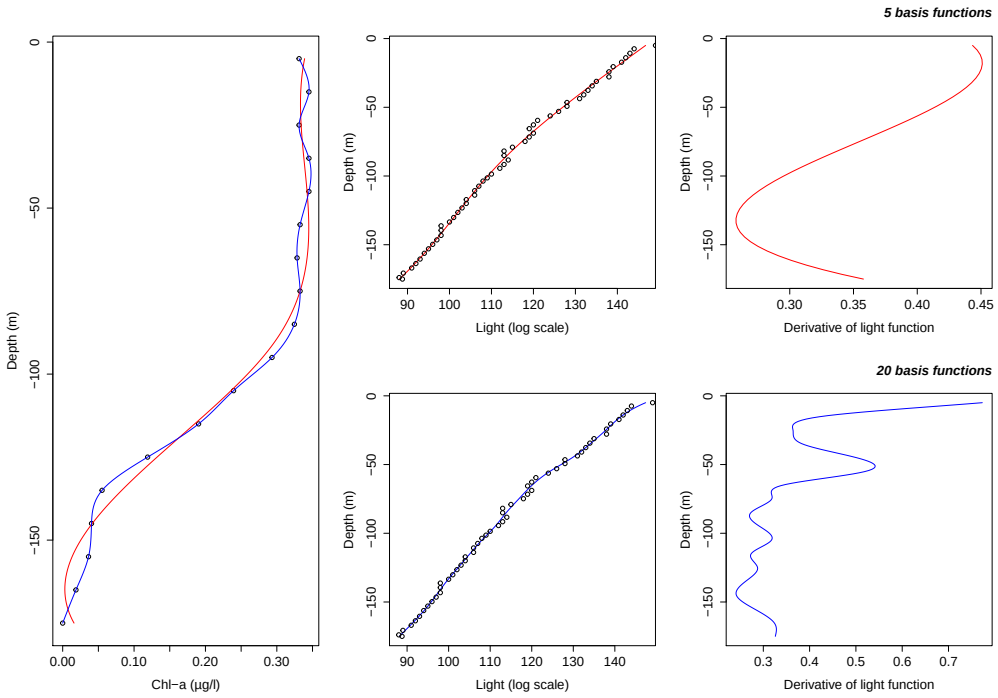


Figure C: Examples of Chl-a and light profiles fitted with 5 or 20 basis functions, with derivatives of light profiles. Red curves show fits with 5 basis functions, and blue curves represent fits with 20 basis functions. The more the basis functions number increase, the more the fit is accurate. For light profiles fits, monotony constraints have been taken into account.

3.3. Prediction errors

Once the dataset of profiles constructed, parameters α and β of model 2 can be estimated following the procedure in appendix A. Considering a light

profile $L(z)$ and the derivative of its logarithm $L'(z)$, the functional linear model gives a Chl a predicted curve denoted by $\widehat{C}$ such that

$$\widehat{C}_i(z) = \widehat{\alpha}(z) + \int_{Z_m}^{Z_M} \widehat{\beta}(s,z) L_i'(s) ds.$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are parameter estimations. The predictive capabilities of the model can be measured with the integrated mean squared error

$$SSE = \frac{1}{n} \sum_{i=1}^{n} \int_{Z_m}^{Z_M} \left(C_i(z) - \widehat{C}_i(z) \right)^2 dz.$$

In the following, we will use expressions derived from this quantity in order to select the right number of basis functions and to test if differences appear between data sampled by different animals.

3.3.1. Choosing the right number of basis functions

Leave-one-out cross validation method is used to find the right number of basis functions for fitting Chl a profiles and light profiles from raw data (Ref. Efron and Tibshirani, 1990). Consider data from a unique elephant seal denoted as e. Consider the observed sample of size n_e $\{(L_i, C_i), i = 1, \ldots, n_e\}$. These profiles have been constructed with K basis functions for the functions C_i and M basis functions for profiles L_i using the raw data. Let $SSE_{CV}(K, M)$ be the mean prediction error integrated alongside depth and defined as

$$SSE_{CV}(K,M) = \frac{1}{n_e} \sum_{i=1}^{n_e} \int_{Z_m}^{Z_M} \left(C_i(z) - \widehat{C}_{(-i)}(z) \right)^2 dz.$$

The curve $\widehat{C}_{(-i)}$ is the prediction of the observed profile C_i when the estimated parameters $\widehat{\beta}_{(-i)}(s,z)$ and $\widehat{\alpha}_{(-i)}(z)$ of the functional linear model have been computed using the whole set of observations except observation (L_i, C_i). The computation of $SSE_{CV}(K, M)$ can be achieved for various values of pair (K, M) giving rise to a graph such as in Fig. D. The right number of basis functions is provided by the pair $(K^\star, L^\star)$ which minimizes the cross-validated prediction error $SSE_{CV}(K, M)$.

3.3.2. Cross validated prediction between animals

Once the number of basis functions has been selected, the whole set of n matched observed profiles is constructed. Using that curves, the parameters of the functional linear model 2 are estimated. This model can then be used to give prediction of a Chl a profile using the corresponding light profile. It can be interesting to assess error variance of prediction between animals. Consider again the elephant seal e and estimate the parameters of a linear model 2 using the n_e profiles $(L_i, C_i), i = 1, \ldots, n_e$. Prediction error can be computed using the $n - n_e$ remaining profiles from other elephant seals such that

$$SSE_e = \frac{1}{(n - n_e)} \sum_{j=1}^{n-n_e} \int_{Z_m}^{Z_M} \left(C_j(z) - \widehat{C}_j^{(e)}(z) \right)^2 dz,$$

where $\widehat{C}_j^{(e)}$ is the predicted Chl a profile using the linear model whose parameters have been estimated with data from elephant seal e.

4. Results

Cross validation indicated globally that only five basis functions are required for both Chl a and light profiles to get a minimum error between the fit and the prediction of Chl a profiles (Fig. D). We applied the functional linear model alternately on 207 light profiles selected to compute at the same time prediction intervals by bootstrap (see paragraph 3.4.).

Several examples of results are represented in Fig. E. The first eight graphics (Fig. E.1 to Fig. E.8) show well predicted Chl a profiles. Black curves (fitted profiles) and red curves (predicted profiles) have the same shape and amounts of Chl a predicted by the functional linear model every depth match fairly well to raw observations.

Two types of prediction problems can be distinguished. Firstly, there are prediction problems at surface or in depth (Fig. E.9 and Fig. E.10). In most cases, but not always, a poor reconstruction of Chl a concentration estimated from light data tends to overestimate Chl a assessed from the light data compared to the fluorescence method. This pattern is likely to be the consequence of greater light attenuation. Particles in suspension in the water column other than phytoplankton can contribute to light attenuation, *e.g.* inorganic particles such as CDOM and also zooplankton. Furthermore, fluorescence in itself is only a proxy of phytoplankton concentration and

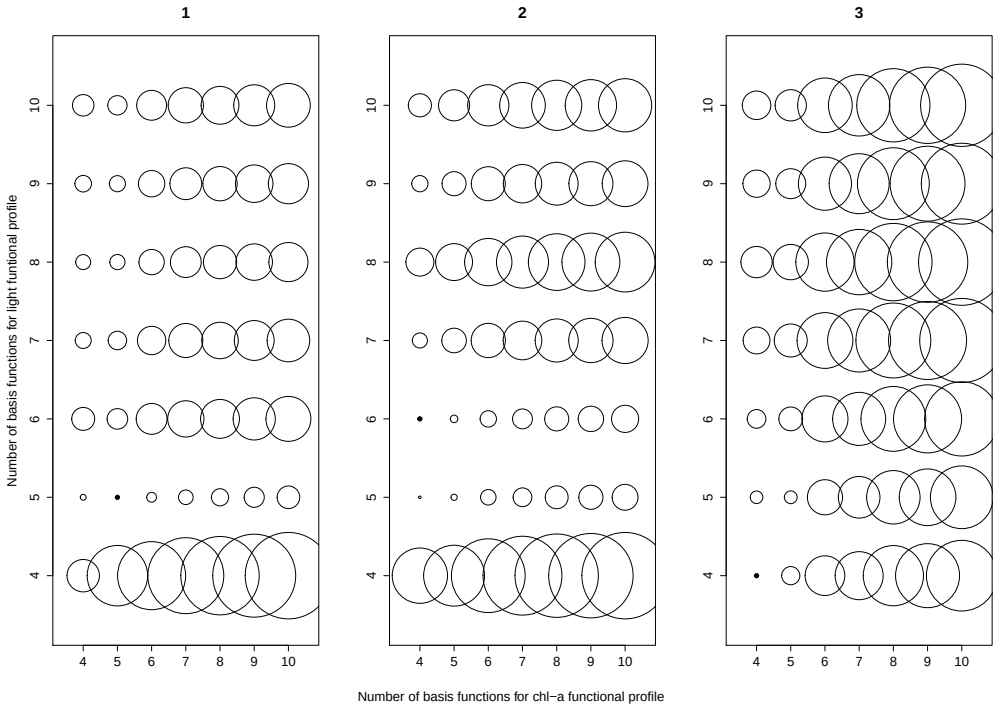


Figure D: Errors between the fit and the prediction of Chl *a* profiles, according to basis function number used to fit Chl *a* and light profiles, for each elephant seal. We can see that five basis functions are required to adjust Chl *a* and light profiles for mammal 1 because minimum error is of value 12 (1). For elephant 2 (2), four basis functions for Chl *a* and six basis functions for light profiles are sufficient. Finally, only four basis functions for each category for elephant 3 are required (3).

fluorescence response is known to vary with phytoplankton species and physiological state (Xing et al., 2012). Phytoplankton species differ widely in their size and shape. Therefore, for a given Chl *a* concentration estimated from the fluorometer, we may expect a variation in the light attenuation factor according to phytoplankton species. A comparison with data of high resolution measurement of ocean colour from PHYSAT which distinguishes the dominant phytoplankton groups within a given area (Alvain et al., 2005) could be used in the future to assess the effect of phytoplankton species on light attenuation. As coastal areas were excluded from our study, we believe that turbidity had a limited effect on the light attenuation coefficients we calculated.

Secondly, profiles can be badly predicted all along depth with a high prediction error. Predicged profiles can be either of very different shape (Fig. E.11) or the whole predicted curve is shifted from the observed profile even if the shape has been kept (Fig. E.12). This last problem refers to an offset problem: when the sensor does not measure Chl a, it returns to 0.

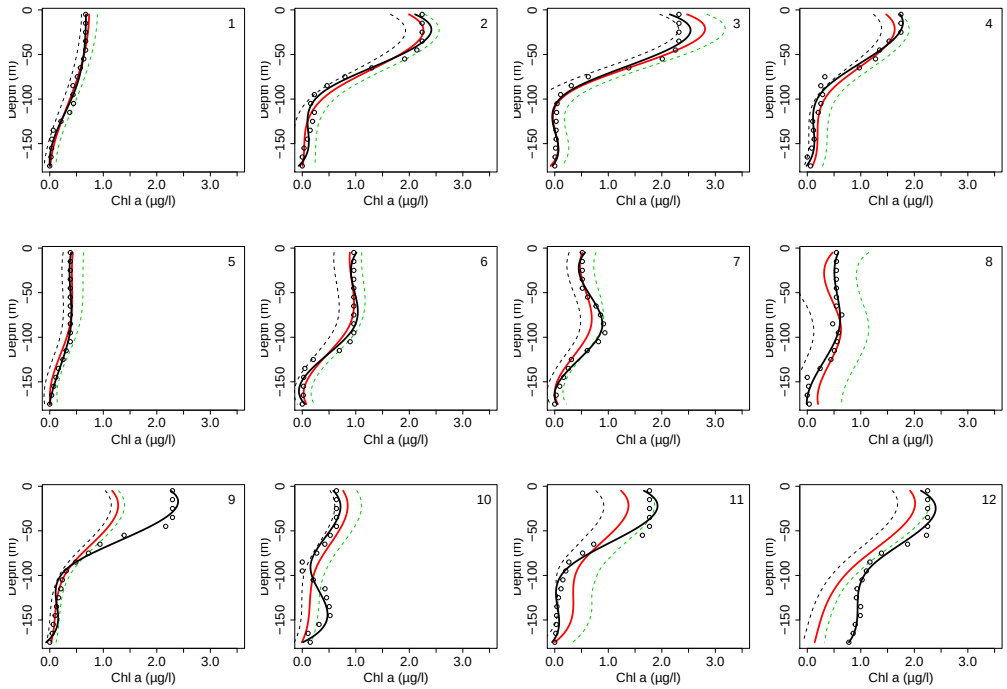


Figure E: Examples of prediction of Chl a profiles. Red curves show profiles reconstructed from measurements, and black curves represent profiles predicted with the fully functional linear model. Profiles 1, 2, 3, 4, 9 and 12 were taken on the path of the elephant seal number 1; profiles 5, 6, 7, 10 and 11 were recorded on the elephant seal number 2, and profile 8 is taken on the number 3.

A measure of goodness of fit can be assessed by constructing a coefficient of determination. For each pair of profiles, define

$$R_i^2 = 1 - \frac{\frac{1}{n}\int \left(C_i(z) - \widehat{C}_i(z)\right)^2 dz}{\frac{1}{n}\int \left(C_i(z) - \overline{C}_i(z)\right)^2 dz}$$

where $C_i(z)$ is a fitted Chl a profile, $\overline{C}(z)$ the mean function of the sample

$\{C_1, \cdots, C_n\}$ defined as

$$\overline{C}(z) = \frac{1}{n} \sum_{i=1}^{n} C_i(z)$$

and $\widehat{C}_i(z)$ the predicted Chl *a* profile for observation $C_i(z)$. This quantity ranges from 0 to 1 and has no unit.

Fig. F displays boxplots of R_i^2 between the predicted and observed Chl *a* profiles for each elephant seal path (1, 2 and 3). Higher values indicate that the model fits the data better. More than 50% of 73 pairs of profiles have a R^2 above 0.71 for seal 1. For seal 2, 50% of 73 pairs of profiles have a R^2 above 0.52. Equally, on the third, 50% of 62 pairs of profiles have a R^2 above 0.52.

However, following the last definition of the R_i^2 negative values can occur if the total variance is smaller than the sum of the predicted squared errors. This indicates that the hypothesis of the functional model do not apply for these observations. Then, eight pairs have negative values (-0.008, -0.07, -0.13, -0.23, -0.43, -1.13, -2.15, -21.77) for path. 1. There are twenty-six pairs of profiles with a negative R^2 with values ranging between -0.09 and -14.96 for path. 2. Finally, path. 3 gave thirteen pairs with a negative R^2 value (-0.0001, -0.11, -0.52, -0.71, -0.06, -0.11, -0.03, -0.08, -1.07, -0.31, -0.17, -0.12, -0.31). These ill-predicted profiles often appears in cases where observations of Chl *a* are close to zero.

The values of the R_i^2 are also affected when prediction error is estimated between animals but the R_i^2 stays globally high for path 1 and path 2 (not presented here).

As the prediction errors for predicted Chl *a* profiles are fairly good, the functional linear model can be used to predict Chl *a* concentration in areas where no amount of Chl *a* was measured. Fig. G.1 and Fig. G.2 show examples of predicted Chl *a* profiles on two different days. Consecutive light fitted Chl *a* profiles reveals a high degree of temporal and therefore spatial variability. The mean travel distance between dives is 1.4 $\pm$0.8 km (range 0.08-5.4 km). Wide variations in Chl *a* concentration are detected at the scale of 2 to 3 dives corresponding to a sub-mesoscale spatial variation (*i.e.* 3 to 5 km) of the phytoplankton. Smaller scales are likely to occur as in a few instances, wide variations in Chl *a* concentration are detected between consecutive dives (Fig. G).

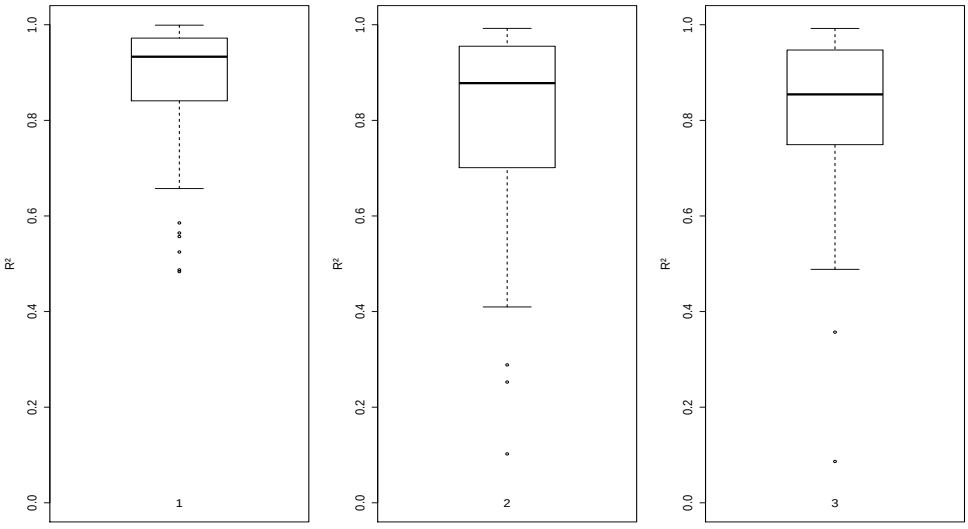
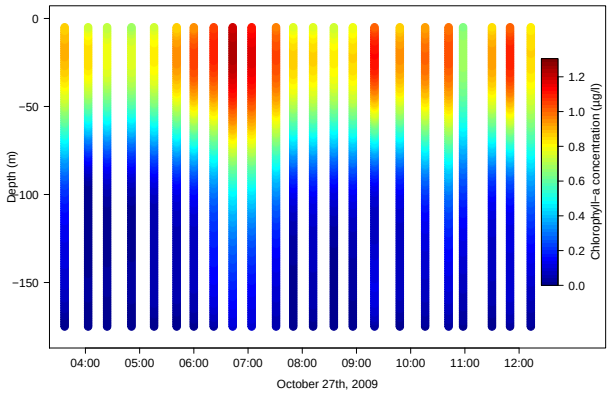


Figure F: Boxplots of R^2 between reconstructed and predicted Chl a profiles for each path of elephant seal (1, 2 and 3). The best goodness of fit concerns the first elephant, with a median R^2 equals to 0.93 (1).

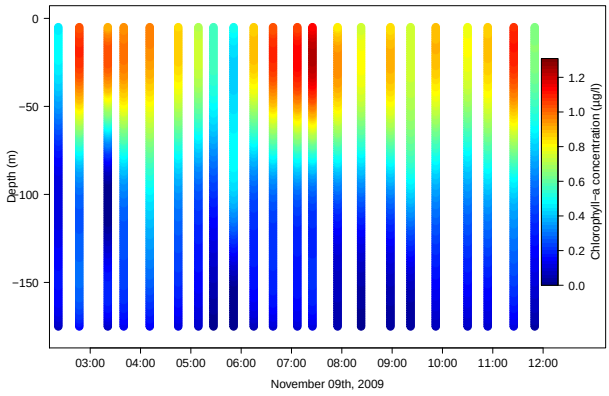
5. Discussion

The first objective of this work was to assess the quality of the predicted Chl a profiles from light data by comparing them with quenching corrected fluorescence profiles (Xing et al., 2012) simultaneously measured by the SRDL CTD-Fluo tags and which were quenching corrected (Guinet et al., 2013b). The second objective was to predict Chl a profiles from light data for each dive occurring during daylight hours throughout the path followed by the Southern elephant seals to infer change in Chl a concentration at sub-mesoscale.

The proposed functional linear model, constructed from physical assumptions, was well suited to predict Chl a profiles from the derivative of log-light profiles and to infer changes in Chl a concentration at fine-scale. This prediction is done for each dive occuring during daylight hours along paths of Southern elephant seals. The reconstruction of Chl a profiles from light measurements (Teo et al., 2009) and more particularly the question of the accuracy of such reconstructions was raised in Jaud et al. (2012). The pro-



1



2

Figure G: Two examples of prediction of Chl *a* concentration in a range between two mesured Chl *a* profiles. This prediction highlights sub-mesoscale variations. The more color is red, the more Chl *a* concentration is high.

cedure is based on (i) a basis function expansion, and (ii) a fully functional linear model. Chl *a* and light profiles are reconstructed as continuous curves using projection in basis functions, and the fully functional linear model can be used to predict functional Chl *a* profiles from derivative of functional light profiles.

This method can be applied under similar conditions in other contexts. To apply our procedure, we need available light measurements recorded between the surface and the depth to which the Chl *a* data interest us, and absorption

coefficients. Data can be recorded with tags fixed on marine mammals or with drifting buoys, for example. We can take into account a light profile as long as the difference of light measurements between the profile surface and the profile deep maximum (180 meters) is sufficient. If this is not the case, prediction of Chl a functional profiles may not be accurate because light data are zero since they are as recorded at night. However, in our study, only Chl a profiles recorded during daylight hours were analyzed. To account for Chl a profiles registered by night, we propose to develop in future an interpolation using kriging methods when data are curves (Nerini et al., 2010).

In this article, we consider a fully functional linear model where both the response C and the explanatory variable L are functions. Teo et al. (2009) used in their study a modified version of the bio-optical model developed by Morel (1988). This bio-optical model allows only the use of data at the same depth $s = z$. With a fully functional linear model, we can predict Chl a concentration at different depths s and z, and at the same time. However, the fully functional linear model takes into account only one explanatory variable, although in theory one can use several. In our case, temperature or salinity profiles could be added in this model to have more information on mixed layers which is of critical importance in controlling the vertical distribution of phytoplankton, and therefore may help in assessing more precisely the depth of rupture point in Chl a concentration within the water column (Boyd, 2002; Chiswell, 2011; Taylor and Ferrari, 2011). So to incorporate other explanatory variables in our model, one have to generalize to the multivariate case.

The method is implemented with parameter value 4 for the order of splines and 5 for number of basis functions. Computation by cross-validation of the optimum number of basis functions to use showed us that the use of only five basis functions is enough to minimize the prediction error. However, this number is optimum in our case, but it can be different to fit other profiles of physical data such as temperature. This highlights the need to choose this number by a calculation method adaptable to any kind of profile. If we increase the order of splines and/or number of basis functions, there is a risk of over-fit curves on data. Furthermore, in order to make the derivative positive, we imposed a monotonic smoothing on functional light profiles. This method doesn't have to be applied to another kind of data, such as temperature or salinity, because it is specific to the decrease of light data when Southern elephant seal goes deep in the water.

Concerning prediction intervals bootstrap, there is still room for improve-

ment. These intervals seem a little tight around a predicted Chl *a* profile. This is due to the fact that prediction was made profile by profile. Furthermore, if basis functions are not correctly adjust data (see Discussion above), intervals do not contain fitted Chl *a* profiles.

The measurement devices and sensors may be subject to bias and measurement errors. In an attempt to correct instrument bias and calibration issues, the data day have been pre-processed before analysis. Data were corrected for the offset (which is a problem of shifting the origin) and for the quenching (a process which decreases the fluorescence intensity). This pretreatment has a relative influence on the Chl *a* profiles adjustment. Some instrument bias corrections can be made before analysis, but for others, there would be very probably an advantage to making them at the same time as the smoothing and adjustment.

In some cases, Chl *a* profiles can be affected by quenching effect. This results from a planktonic reaction which disturbs the fluorescence measurements. Light data could help to correct fluorescence profiles affected by quenching effect, especially in stratified waters, where this cannot be achieved by profile-by-profile quenching correction (Xing et al., 2012). In fact, light attenuation data are only dependent on the presence of phytoplanktonic particles, not dependent on their physiological state and thus are not affected by the quenching effect.

In-situ high spatial resolution of vertical profiles of phytoplankton concentration provided by the Southern elephant seals from light attenuation profiles combined with remotely sensed ocean colour and sea surface temperature images represent a significant contribution in assessing the (sub-)mesoscale spatial structuration of areas of ecological importance within the Kerguelen Region. The simultaneous collection of information on Southern elephant seals foraging success assessed from head mounted accelerometers (Gallon et al., 2012; Guinet et al., 2013a) with concomitant fine scale oceanographic variables including the spatial distribution of phytoplankton should provide new insight into the spatial structuring of the prey field in relation to the oceanographic landscape at (sub-)mesoscale.

6. Acknowledgments

The authors wish to thank the Provence-Alpes-Côte-d'Azur Regional Council for funding. This work was funded by the ANR project IPSOS-SEAL (French National Research Agency). The authors also thank IPEV

and Total Foundation.

Alvain, S., Moulin, C., Dandonneau, Y., Bréon, F., 2005. Remote sensing of phytoplankton groups in case 1 waters from global seawifs imagery. Deep-Sea Research I 52, 1989–2004.

Arrigo, K., Worthen, D., Lizotte, M., Dixon, P., Dieckmann, G., 1997. Primary production in antarctic sea ice. Science 276 (5311), 394–397.

Behrenfeld, M., Falkowski, P., 1997. A customer's guide to phytoplankton primary productivity models. Limnology and Oceanography 42 (7), 1479–1491.

Behrenfeld, M., O'Malley, R., Siegel, D., McClain, C., Sarmiento, J., Feldman, G., Milligan, A., Falkowski, P., Letelier, R., Boss, E., 2006. Climate-driven trends in contemporary ocean productivity. Nature 444, 752–755.

Biuw, M., Boehme, L., Guinet, C., Hindell, M., Costa, D., Charrassin, J.-B., Roquet, F., Bailleul, F., Meredith, M., Thorpe, S., Tremblay, Y., McDonald, B., Park, Y.-H., Rintoul, S., Bindoff, N., Goebel, M., Crocker, D., Lovell, P., Nicholson, J., Monks, F., Fedak, M., 2007. Variations in behavior and condition of a southern ocean top predator in relation to in situ oceanographic conditions. Proceedings of the National Academy of Sciences 104 (34), 13705–13710.

Block, B., Costa, D., Boehlert, G., Kochevar, R., 2002. Revealing pelagic habitat use: the tagging of pacific pelagics program. Oceanologica Acta 25 (5), 255–266.

Boehlert, G., Costa, D., Crocker, D., Green, P., O'Brien, T., Levitus, S., Le Boeuf, B., 2001. Autonomous pinniped environmental samplers: using instrumented animals as oceanographic data collectors. Journal of atmospheric and oceanic technology 18 (11), 1882–1893.

Boehme, L., Lovell, P., Biuw, M., Roquet, F., Nicholson, J., Thorpe, S., Meredith, M., Fedak, M., 2009. Technical note: Animal-borne ctd-satellite relay data loggers for real-time oceanographic data collection. Ocean Science 5 (4), 685–695.

Bouguer, P., 1729. Essai d'optique sur la gradation de la lumière. Claude Jombert, Paris.

Boyd, P., 2002. Environmental factors controlling phytoplankton processes in the southern ocean. Journal of Phycology 38 (5), 844–861.

Buesseler, K., Boyd, P., 2003. Will ocean fertilization work? Science 300 (5616), 67–68.

Charrassin, J., Hindell, M., Rintoul, S. R., Roquet, F., Sokolov, S., Biuw, M., Costa, D., Boehme, L., Lovell, P., Coleman, R., Timmermann, R., Meijers, A., Meredith, M., Park, Y.-H., Bailleul, F., Goebel, M., Tremblay, Y., Bost, C.-A., McMahon, C. R., Field, I. C., Fedak, M. A., Guinet, C., 2008. Southern ocean frontal structure and sea-ice formation rates revealed by elephant seals. Proceedings of the National Academy of Sciences 105 (33), 11634–11639.

Chiswell, S., 2011. Annual cycles and spring blooms in phytoplankton: don't abandon sverdrup completely. Marine ecology progress series 443, 39–50.

Craven, P., Wahba, G., 1979. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numerische Mathematik 31, 377–403.

Fedak, M., 2013. The impact of animal platforms on polar ocean observation. Deep Sea Research Part II: Topical Studies in Oceanography 88, 7–13.

Fedak, M., Lovell, P., McConnell, B., Hunter, C., 2002. Overcoming the constraints of long range radio telemetry from animals: getting more useful data from smaller packages. Integrative and Comparative Biology 42 (1), 3–10.

Gallon, S., Bailleul, F., Charrassin, J.-B., Guinet, C., Bost, C.-A., Handrich, Y., Hindell, M., 2012. Identifying foraging events in deep diving southern elephant seals, mirounga leonina, using acceleration data loggers. Deep Sea Research Part II: Topical Studies in Oceanography.

Guinet, C., Vacquié-Garcia, J., Picard, B., Bessigneul, G., Lebras, Y., Dragon, A.-C., Viviant, M., Arnould, J., Bailleul, F., 2013a. Southern elephant seal foraging success in relation to temperature and light conditions: insight on their prey distribution, unpublished results.

Guinet, C., Xing, X., Walker, E., Monestiez, P., Marchand, S., Picard, B., Jaud, T., Authier, M., Cotté, C., Dragon, A.-C., Diamond, E., Antoine,

D., Lovell, P., Blain, S., D'Ortenzio, F., Claustre, H., 2013b. Calibration procedures and first data set of southern ocean chlorophyll-a profiles collected by elephant seal equipped with a newly developed ctd-fluorescence tags. Earth System Science Data 5, 15–29.

He, G., Müller, H.-G., Wang, J.-L., Yang, W., 2010. Functional linear regression via canonical analysis. Bernoulli 16 (3), 705–729.

Holm-Hansen, O., Amos, A., Hewes, C., 2000. Reliability of estimating chlorophyll a concentrations in antarctic waters by measurement of in situ chlorophyll a fluorescence. Marine Ecology Progress Series 1996, 103–110.

Hosseini-Nasab, M., 2012. Cross-validation approximation in functional linear regression. Journal of Statistical Computation and Simulation, 1–11. URL http://www.tandfonline.com/doi/abs/10.1080/00949655.2012.662502

Jaud, T., Dragon, A.-C., Garcia, J., Guinet, C., 2012. Relationship between chlorophyll a concentration, light attenuation and diving depth of the southern elephant seal mirounga leonina. PLoS one 7 (10), e47444.

Kiefer, D., 1973. Fluorescence properties of natural phytoplankton populations. Marine Biology 22, 263–269.

Marra, J., 1997. Analysis of diel variability in chlorophyll fluorescence. Journal of Marine Research 55, 767–784.

Maxwell, K., Johnson, G., 2000. Chlorophyll fluorescence - a practical guide. Journal of experimental botany 51 (345), 659–668.

McMahon, C. R., Autret, E., Houghton, J., Lovell, P., Myers, A., Hays, G., 2005. Animal-borne sensors successfully capture the real-time thermal properties of ocean basins. Limnology and Oceanography: Methods 3, 392–398.

Morel, A., 1988. Optical modeling of the upper ocean in relation to its biogenous matter content (case i waters). Journal of Geophysical Research: Oceans (1978–2012) 93 (C9), 10749–10768.

Nerini, D., Monestiez, P., Manté, C., 2010. Cokriging for spatial functional data. Journal of Multivariate Analysis 101, 409–418.

Queguiner, B., Brzezinski, M., 2002. Biogenic silica production rates and particulate organic matter distribution in the atlantic sector of the southern ocean during austral spring 1992. Deep Sea Research Part II: Topical Studies in Oceanography 49, 1765–1786.

Ramsay, J., Silverman, B., 2005. Functional Data Analysis. New-York.

Taylor, J. R., Ferrari, R., 2011. Shutdown of turbulent convection as a new criterion for the onset of spring phytoplankton blooms. Limnology and Oceanography 56 (6), 2293–2307.

Teo, S., Kudela, R., Rais, A., Perle, C., Costa, D., Block, B., 2009. Estimating chlorophyll profiles from electronic tags deployed on pelagic animals. Aquatic Biology 5, 195–207.

Vacquié-Garcia, J., 2014. Variation spatio-temporelle de l'activité d'alimentation des éléphants de mer en relation avec les paramètres physiques et biologiques de l'environnement. Ph.D. thesis, Université Paul Sabatier.

Xing, X., Claustre, H., Blain, S., F., D., Antoine, D., Ras, J., Guinet, C., 2012. Quenching correction for in vivo chlorophyll fluorescence measured by instrumented elephant seals in the kerguelen region (southern ocean). Limnology and Oceanography: Method 10, 483–495.

Appendix A. Parameters estimation for functional linear model with functional covariate

Suppose that we dispose of n pairs $\{(L_i, C_i), i = 1, \ldots, n\}$ which define a set of matched curves sampled from functional variables L and C. We suppose that these functions belong to $\mathbb{L}^2[0; Z]$ the space of square integrable functions defined on the bounded interval $[0; Z]$. This functional space is equipped with the inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. We consider a functional regression model in which the function $L(s)$, $s \in [0, Z]$ is used as a covariate to explain the variation of the response curve $C(z)$, $z \in [0, Z]$. The most general version of a functional linear model using a functional covariate is given by

$$C = \alpha + B(L) + \varepsilon$$

where the intercept $\alpha(z)$ is a functional parameter and B is a linear operator such that

$$B(L)(z) = \int_0^Z \beta(s, z) L(s) \, ds.$$

The kernel of the operator B is a bivariate function $\beta(s, z)$ that acts as a regression coefficient. It potentially gives the influence of $L(s)$ on $C(z)$ at any value of z. The functional remainder ε gives the error between the model and the function C. The coefficients α and β must be estimated using the sample at hands.

The search for estimates of $\widehat{\alpha}$ and $\widehat{B}$ is achieved when minimizing the expectation of the quadratic error

$$SSE(\alpha, B) = \mathbb{E}\left(\|\varepsilon\|^2\right).$$

The solution of that minimization problem leads to the normal equations for the functional linear model

$$\left\{ \begin{array}{rcl} V_L B & = & V_{LC} \\ \alpha & = & \mu_C - B(\mu_L) \end{array} \right. \tag{A.1}$$

where functions μ_L and μ_C are expectations of variables L and C respectively, V_L is the variance-covariance operator for variable L and V_{LC} is the cross-covariance operator between variables L and C. Empirical versions of estimators for both covariance operators and for mean functions are computed from the sample $\{(L_i, C_i), i = 1, \ldots, n\}$ as

$$\left\{\begin{array}{rcl} \widehat{V}_L & = & \frac{1}{n}\sum_{i=1}^{n}\left(L_i-\widehat{\mu}_L\right)\otimes\left(L_i-\widehat{\mu}_L\right) \\ \widehat{V}_{LC} & = & \frac{1}{n}\sum_{i=1}^{n}\left(L_i-\widehat{\mu}_L\right)\otimes\left(C_i-\widehat{\mu}_C\right) \\ \widehat{\mu}_L & = & \frac{1}{n}\sum_{i=1}^{n}L_i \\ \widehat{\mu}_C & = & \frac{1}{n}\sum_{i=1}^{n}C_i \end{array}\right.$$

where the tensor product of two elements X and Y of $\mathbb{L}^2[0;Z]$ is the rank one operator such that $[X\otimes Y](f)=\langle X,f\rangle Y$ for all $f\in\mathbb{L}^2[0;Z]$. The straightforward estimators $\widehat{B}$ of B and $\widehat{\alpha}$ of α are obtained by replacing empirical estimators in (A.1) such that

$$\left\{\begin{array}{rcl} \widehat{V}_L\widehat{B} & = & \widehat{V}_{LC} \\ \widehat{\alpha} & = & \widehat{\mu}_C-\widehat{B}\left(\widehat{\mu}_L\right) \end{array}\right. .$$

One way to achieve the above calculus of is to decompose both $L(z)$ and $C(z)$ as a linear combination of known basis functions. With this decomposition, the problem of estimation, including the calculus of the inverse of $\widehat{V}_L$, can then be handled as a known multivariate problem, working on the coefficients of the basis decomposition (He et al., 2010). [inclure plus de détails].

Appendix B. Bootstrap pointwise prediction intervals

Once the parameters have been estimated, the functional linear model gives a predicted value of the response curve at depth z such that

$$\widehat{C}_i(z)=\widehat{\alpha}(z)+\int_0^Z\widehat{\beta}(s,z)\,L_i(s)\,ds$$

where $\widehat{\alpha}$ and $\widehat{\beta}$ are the functional parameters estimated using the initial sample $\{(L_i,C_i),i=1,\ldots,n\}$ constructed with their right number of basis functions. Denote $\widehat{\varepsilon}_i(z)=C_i(z)-\widehat{C}_i(z)$ the estimated residuals at fixed depth z. The forecast of an independent observation $C_{n+1}(z)$ is then

$$\widehat{C}_{n+1}(z)=\widehat{\alpha}(z)+\int_0^Z\widehat{\beta}(s,z)\,L_{n+1}(s)\,ds.$$

A bootstrap replication of the initial observations and a future value are given by the pairs $\{(L_i,C_i^\star),i=1,\ldots,n+1\}$ such that

$$C_i^\star(z)=\widehat{\alpha}(z)+\int_0^Z\widehat{\beta}(s,z)\,L_i(s)\,ds+\varepsilon_i^\star(z)$$

for $i = 1, \ldots, n$ and

$$C^{\star}_{n+1}(z) = \widehat{\alpha}(z) + \int_0^Z \widehat{\beta}(s,z)\, L_{n+1}(s)\, ds + \varepsilon^{\star}_{n+1}(z)$$

where $\varepsilon^{\star}_1(z), \ldots, \varepsilon^{\star}_n(z)$ and $\varepsilon^{\star}_{n+1}(z)$ are obtained by sampling with replacement from the empirical distribution of the pointwise residuals

$$\widehat{F}_z(x) = \#\left\{\widehat{\varepsilon}_i(x) \leq z\right\}/n.$$

Let $\widehat{\alpha}^{\star}$ and $\widehat{\beta}^{\star}$ be the estimated parameters using the bootstrap sample $B = \{(L_i, C^{\star}_i), i = 1, \ldots, n\}$ and define the prediction error

$$e^{\star}_{n+1}(z) = C^{\star}_{n+1}(z) - \left[\widehat{\alpha}^{\star}(z) + \int_0^Z \widehat{\beta}^{\star}(s,z)\, X_{n+1}(s)\, ds\right]$$

which has distribution $G^{\star}_t(\cdot; n+1)$. The pointwise bootstrap γ-prediction interval of observation $Y_{n+1}(z)$ is given with

$$I^{(B)}_{\gamma}(z; n+1) = \left[\widehat{C}_{n+1}(z) + G^{\star -1}_z\left((1-\gamma)/2; n+1\right), \widehat{C}_{n+1}(z) + G^{\star -1}_z\left((1+\gamma)/2; n+1\right)\right]$$

where $G^{\star -1}_z(\cdot; n+1)$ is the quantile function that gives the $(1-\gamma)/2$ lower quantile and the $(1+\gamma)/2$ upper quantile $(0 \leq \gamma \leq 1)$.

The distribution of the prediction errors is then simulated using bootstrapped errors obtained by sampling the empirical distribution function $\widehat{F}_z$ in place of F_z and using $\widehat{\alpha}$ and $\widehat{\beta}$ as true coefficients of the regression instead of α and β. The estimation of a prediction interval can then be repeated for any value of depth z to form the functional prediction envelope (Fig. C).

In our case, the covariate is the derivative of log-L.