

# Des données satellites de couleur de l'eau aux propriétés biogéochimiques

Master d'Océanographie II- Spécialité OPB  
OPB305 Optique Marine

Alain Fumenia  
Laboratoire d'Océanologie et de Géoscience (LOG)  
[alain.fumenia@univ-littoral.fr](mailto:alain.fumenia@univ-littoral.fr)

## Objectif

Comprendre les principes permettant le développement et la validation d'algorithmes à partir de données satellites pour estimer les variables biogéochimiques de l'océan.

## Sommaire

### 1) Introduction

- 1.1) Les données satellites de couleur de l'eau
- 1.2) Les différents produits couleur de l'océan disponibles
- 1.3) Algorithmes bio-optiques
  - 1.3.1) Algorithmes empiriques
  - 1.3.2) Algorithmes semi-analytiques

### 2) Procédure de développement et validation d'un algorithme

### 3) Cas d'étude : le carbone organique particulaire

- 3.1) Le POC océanique
- 3.2) Première estimation du POC (Stramski et al., 1999)
- 3.3) Estimation du POC dans l'océan ouvert (Stramski et al., 2008)
- 3.4) Validation du POC dans l'océan ouvert
- 3.5) Estimation du POC dans les régions côtières (Kien et al., 2019)
- 3.6) Algorithme adaptatif (Stramski et al., 2023)

### 4) Résumé

# 1) Introduction 1.1) Les données satellites de couleur de l'eau

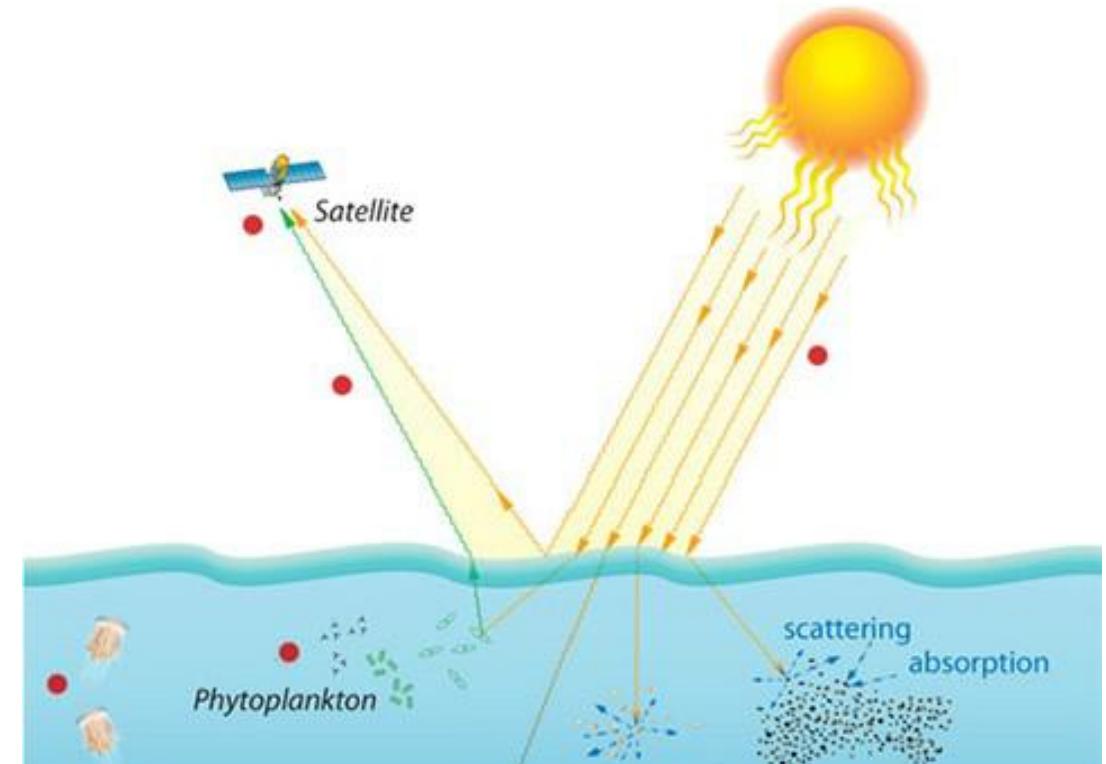
Pour les océanographes, la télédétection permet par **l'acquisition d'image de couleur de l'eau** d'obtenir des **informations sur la surface océanique** sans contact avec cette dernière. Cela nécessite de capter et d'enregistrer l'énergie d'un rayonnement électromagnétique émis ou réfléchis, puis de traiter et mettre en application l'information qu'il représente.

La mesure de la couleur de l'eau correspond à la **mesure du rayonnement solaire réfléchi par l'océan**. Les données sont collectées selon différentes bandes spectrales du domaine visible (400 nm - 700nm) et du proche infra-rouge.

Les satellites sont utilisés pour observer l'océan depuis les années 1970.

Le type d'instrument de mesure de la lumière est un **spectroradiomètre**.

Ce dernier est capable de mesurer la longueur d'onde et l'amplitude de la lumière émise par une source lumineuse.



**Figure 1.** Représentation schématique du devenir de la lumière dans l'océan

# 1) Introduction 1.1) Les données satellites de couleur de l'eau

Satellites	Instruments
• SeaStar	Sea-viewing Wide Field-of-view ( <b>SeaWiFS</b> ; 1997-2010)
• ENVISAT	MEdium Resolution Imaging Spectrometer ( <b>MERIS</b> ; 2002-2012)
• AQUA	Moderate-Resolution Imaging Spectroradiometer ( <b>MODIS-Aqua</b> ; 2002-present)
• SNPP	Visible Infrared Imaging Radiometer Suite ( <b>VIIRS</b> ; 2012-present)
• Sentinel-3 A/B	Ocean and Land Colour Instrument-A and B ( <b>OLCI</b> ; 2016-present)
• Merged product	Produit composite (1997-present)

**Rappel**, la réflectance  $R_{rs}(\lambda)$

$$R_{rs}(\lambda) = \frac{L_w(\lambda)}{E_d(\lambda)} = \frac{\text{Rayonnement ascendant quittant la surface océanique}}{\text{Irradiance descendante arrivant sur la surface océanique}} = G(\lambda) \times \left( \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \right)$$



Méthode **complémentaire** aux mesures in situ.

**Intérêt** : la télédétection permet d'observer le système océanique avec une **haute résolution spatiale** (échelle globale et régionale) et **temporelle** (échelle journalière, mensuelle, annuelle, interannuelle...).

## 1) Introduction 1.2) Les différents produits couleur de l'océan disponibles

Parmi les différents paramètres pouvant être observés depuis l'espace, la mesure de la couleur de l'eau permet d'acquérir des **informations qualitatives et quantitatives** sur la **composition biogéochimique** des eaux de surface.

**Question** : Quelles sont les variables biogéochimiques que l'on peut estimer à partir de l'espace ?

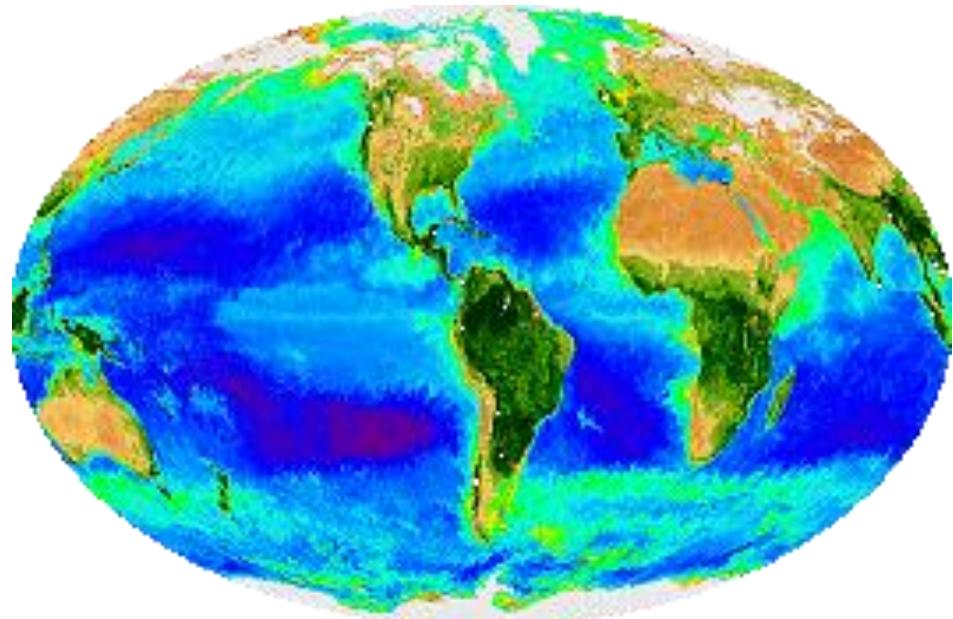
# 1) Introduction 1.2) Les différents produits couleur de l'océan disponibles

Parmi les différents paramètres pouvant être observés depuis l'espace, la mesure de la couleur de l'eau permet d'acquérir des **informations qualitatives et quantitatives** sur la **composition biogéochimique** des eaux de surface.

**Question : Quelles sont les variables biogéochimiques que l'on peut estimer à partir de l'espace ?**

- Chlorophylle a = indicateur de biomasse phytoplanctonique

Les variations de chlorophylle-a sont liées aux variations de l'abondance des organismes phytoplanctoniques



Les variations temporelles de chlorophylle-a reflètent également un ajustement du contenu intracellulaire en chlorophylle-a en fonction de l'éclairement

**Figure 2.** Concentrations en chlorophylle-a à partir d'image de couleur de l'eau du capteur SeaWiFS.

# 1) Introduction 1.2) Les différents produits couleur de l'océan disponibles

**Question : Quelles sont les variables biogéochimiques que l'on peut estimer à partir de l'espace ?**

- Carbone organique particulaire (POC)
  - Carbone organique phytoplanctonique
  - Azote organique particulaire (PON)
  - Carbone organique dissous (DOC)
- } = indicateurs de la biomasse totale (phytoplancton, zooplancton, bactérie, virus, détritus)
- Carbone inorganique particulaire (PIC)
  - Matière particulaire en suspension (SPM) = matière particulaire minérale + organique
  - Production primaire nette
  - Types/groupes phytoplanctoniques
  - Taille des cellules (PSD)

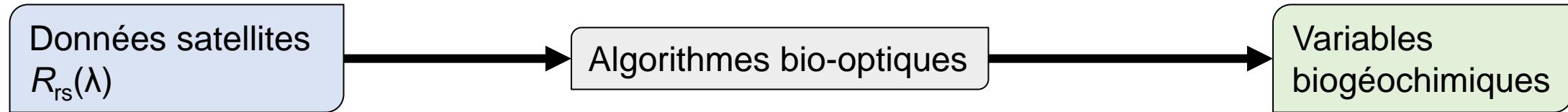
**Résolution spatiale** : océan global

**Résolution temporelle** : 1997-aujourd'hui  
journalière/ 8 jours / mensuelles

# 1) Introduction 1.3) Algorithmes bio-optiques



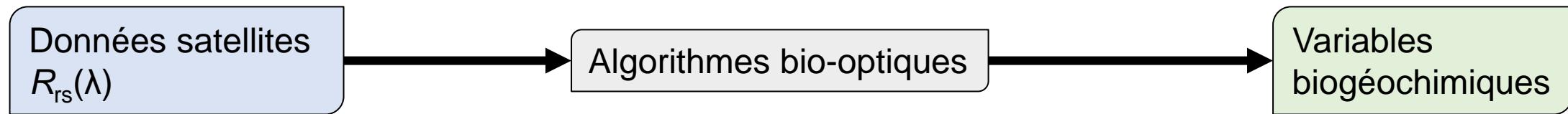
**Algorithme** = une **suite d'instructions logiques et structurées**, permettant de **transformer des données d'entrée** (données optiques) en **informations utiles** (variables biogéochimiques)  
(IOCCG, 2010 ; NASA OBPG)



# 1) Introduction 1.3) Algorithmes bio-optiques



**Algorithme** = une **suite d'instructions logiques et structurées**, permettant de **transformer des données d'entrée** (données optiques) en **informations utiles** (variables biogéochimiques)  
(IOCCG, 2010 ; NASA OBPG)



Pour chaque variable biogéochimique, il existe la plupart du temps différentes méthodes, qui dépendent de l'**objectif scientifique**, de la **région d'intérêt**, des **données disponibles** permettant le développement de l'algorithme



**Challenge** : trouver la **meilleure relation** qui peut exister entre les propriétés optiques mesurées et le paramètre que l'on souhaite estimer

On distingue les algorithmes dit « **empiriques** » et « **semi-analytiques** ».

# 1) Introduction

## 1.3) Algorithmes bio-optiques

### 1.3.1) Algorithmes empiriques

Ce type d'approche s'appuie sur une **régression statistique** (linéaire, polynomiale...) et des **corrélations** reliant une grandeur mesurée ( $R_{rs}$ , IOPs) et une variable biogéochimique.

Le développement d'algorithmes empiriques implique de chercher une combinaison de signaux de  $R_{rs}(\lambda)$  à plusieurs longueurs d'onde afin de trouver un rapport (ex.  $R_{rs}(\lambda_1) / R_{rs}(\lambda_2)$ ) empiriquement lié au paramètre recherché.

# 1) Introduction

## 1.3) Algorithmes bio-optiques

### 1.3.1) Algorithmes empiriques

Ce type d'approche s'appuie sur une **régression statistique** (linéaire, polynomiale...) et des **corrélations** reliant une grandeur mesurée ( $R_{rs}$ , IOPs) et une variable biogéochimique.

Le développement d'algorithmes empiriques implique de chercher une combinaison de signaux de  $R_{rs}(\lambda)$  à plusieurs longueurs d'onde afin de trouver un rapport (ex.  $R_{rs}(\lambda_1) / R_{rs}(\lambda_2)$ ) empiriquement lié au paramètre recherché.

**Exemple** : Concentrations en Chlorophylle-a (Gordon et al., 1983; O'Reilly et al. 1998)

$$\log_{10}(Chla) = a_0 + \sum_{i=1}^N a_i \left( \frac{R_{rs}(\lambda \text{ bleu})}{R_{rs}(\lambda \text{ vert})} \right)^i$$

**Exemple** : Concentrations en POC (Stramski et al., 2008)

$$POC = A \times \frac{R_{rs}(\lambda \text{ bleu})^B}{R_{rs}(\lambda \text{ vert})}$$



L'utilisation des rapports de bandes permet de réduire les artefacts associés aux reflets solaires résiduels, à la lumière parasite et aux erreurs de correction atmosphérique (Hu et al. 2012).



## Hypothèse

Tous les constituants optiques covarient avec le paramètre d'intérêt

### Points positifs :

- Relativement **faciles** à mettre en œuvre.
- Les algorithmes empiriques **sont performants** lorsqu'ils sont appliqués à **l'échelle régionale** en accord avec le jeu de données avec lequel il a été développé.

### Points négatifs :

- Une fois que la relation est établie avec un large jeu de données, il **difficile d'améliorer ces algorithmes**.
- On n'est pas capable de distinguer **l'influence des constituants optiques** qui ne covarient pas avec le paramètre d'intérêt,
- La formulation et la performance peut **dépendre significativement de plusieurs facteurs** (taille et précision du jeu de données in situ, couverture spatiale et saisonnière des mesures, changements environnementaux...)

# 1) Introduction 1.3) Algorithmes bio-optiques

## 1.3.2) Algorithmes semi-analytiques

### Rappel sur les IOPs

Total backscattering coefficient,  $b_b(\lambda)$

$$b_b(\lambda) = b_w(\lambda) + b_{bp}(\lambda)$$

$b_w(\lambda)$  = absorption by water

$b_{bp}(\lambda)$  = backscattering coefficient

Total absorption coefficient  $a(\lambda)$

$$a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{NAP}(\lambda) + a_g(\lambda)$$

$$a_{nw}(\lambda) = a_{ph}(\lambda) + a_{NAP}(\lambda) + a_g(\lambda)$$

$$a_{cdm}(\lambda) = a_{NAP}(\lambda) + a_g(\lambda)$$

$a_w(\lambda)$  = absorption by water

$a_{ph}(\lambda)$  = absorption by phytoplankton

$a_{NAP}(\lambda)$  = absorption by non algal particles

$a_g(\lambda)$  = absorption by CDOM

Au premier ordre, la variabilité des IOPs, comme les coefficients de rétrodiffusion particulaire spectrale,  $b_{bp}(\lambda)$ , et d'absorption particulaire spectrale,  $a_p(\lambda)$ , est principalement déterminée par la **concentration totale de matières particulières en suspension** et par sa **composition** et sa **distribution de taille** au deuxième ordre.



Site utile pour les révisions...

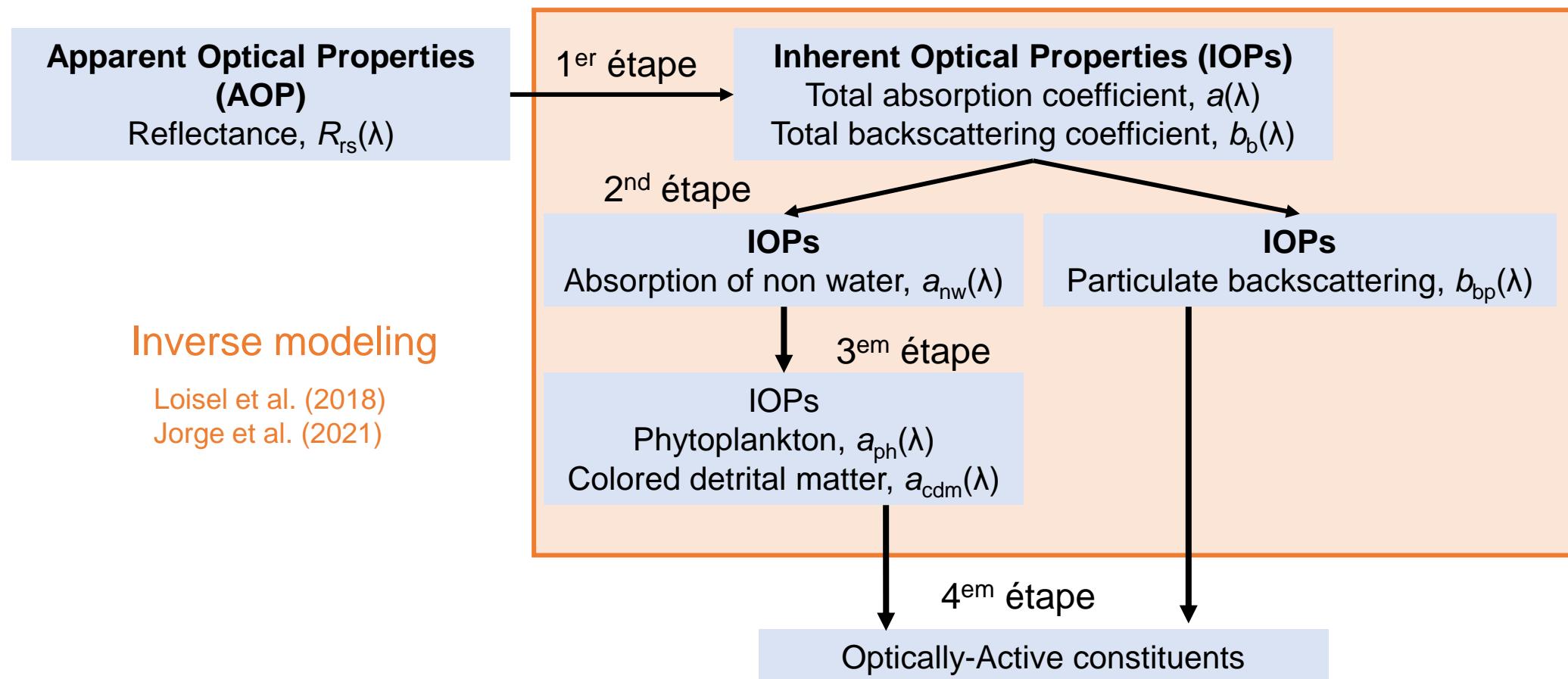
<https://www.oceanopticsbook.info/>

# 1) Introduction

## 1.3) Algorithmes bio-optiques

### 1.3.2) Algorithmes semi-analytiques

Ces algorithmes sont basés sur les résultats de **modèles de propriétés optiques**. Ces modèles sont basés essentiellement sur le lien analytique entre les  $R_{rs}(\lambda)$  et les IOPs établi à partir de la théorie de transfert radiatif (Gordon et al., 1975; Morel and Prieur, 1977).



## Algorithme semi-analytique (ou semi-empirique)



### Points positifs :

- Permet d'estimer **directement plusieurs IOPs** ( $a$ ,  $b_b$ ,  $b_{bp}$ ,  $a_{ph}$ ...) puis de dériver plusieurs variables biogéochimiques.
- Prend en compte **l'interaction lumière-matière**, donc ces algorithmes sont plus cohérents d'un point de vue physique que les algorithmes empiriques.
- Peut être appliqué dans des **eaux du large** (cas 1) mais aussi dans des **zones optiquement complexes** (cas 2, estuaires, zones côtières...)

### Points négatifs :

- Formules plus longues, étapes d'inversion et d'optimisation nécessaires.
- La performance de l'algorithme va dépendre de la qualité des variables optiques dérivées à partir des données satellites.
- Temps de calcul plus couteux que les formules empiriques.
- L'utilisateur doit maîtriser les notions de transfert radiatif.

**Pour développer et valider un algorithme, il est indispensable de suivre différentes étapes :**

- ✓ Base de données in situ
- ✓ Contrôle qualité
- ✓ Base de données de *matchups*
- ✓ Choix du type de modèle de régression
- ✓ Test de performance et validation de l'algorithme

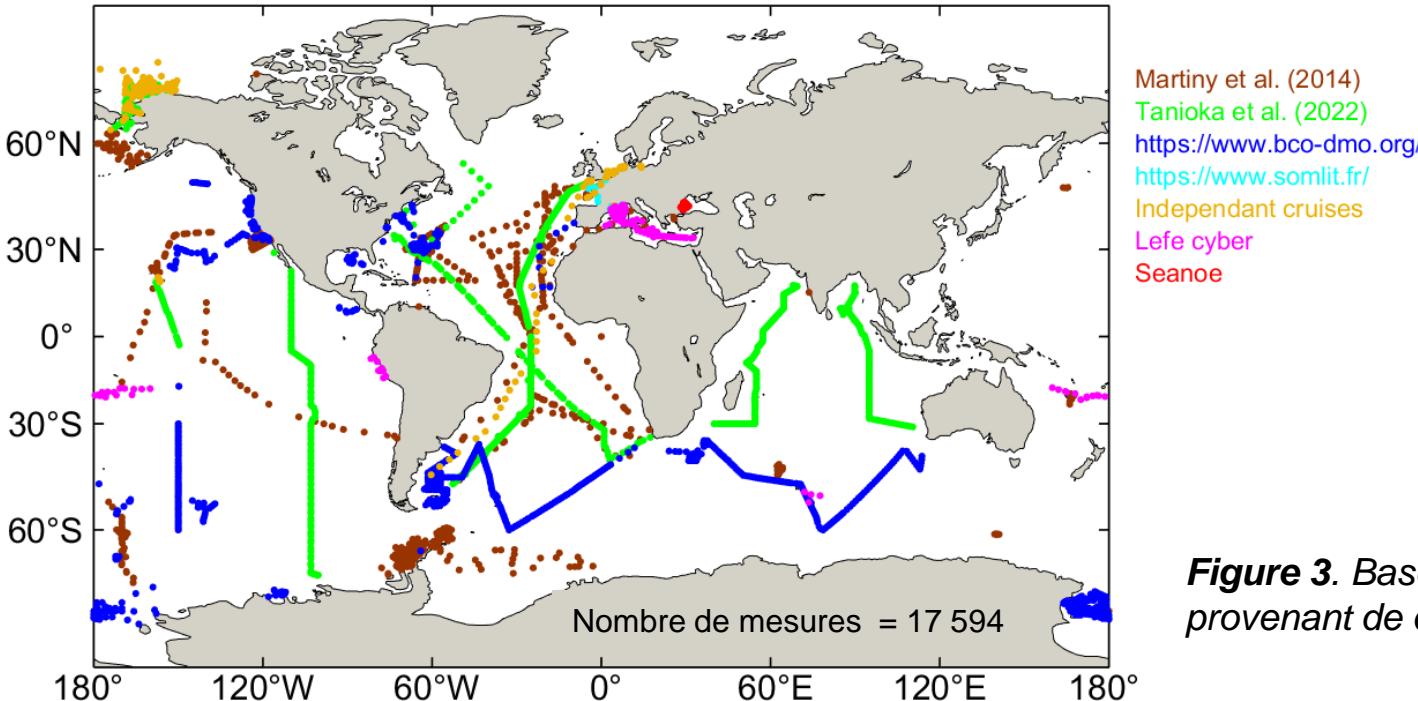
## 2) Procédure de développement et validation d'un algorithme

### Etape 1 : Assembler une base de données *in situ* du paramètre biogéochimique que l'on cherche à estimer



Les algorithmes peuvent être développés à partir d'un jeu de données représentatif de :

- L'océan ouvert uniquement (eaux du cas 1)
- Les régions côtières uniquement (eaux du cas 2)
- L'océan ouvert + régions côtières
- Une région bien spécifique



Couverture temporelle :  
septembre 1997-maintenant

**Figure 3.** Base de données de mesures *in situ* de POC et de PON provenant de différentes bases de données en accès libre.

### Etape 2 : Faire un contrôle qualité rigoureux du jeu de données *in situ*

## 2) Procédure de développement et validation d'un algorithme

**Etape 3** : Pour développer et valider une relation entre une propriété optique et une variable biogéochimique, il est nécessaire de créer une base de données de *matchups*.



La base de données de *matchups* peut être composée soit de :

- Données optiques ( $R_{rs}$ , IOPs) **in situ** + données biogéochimiques **in situ**
- Données optiques ( $R_{rs}$ , IOPs) **satellite** + données biogéochimiques **in situ**

## 2) Procédure de développement et validation d'un algorithme

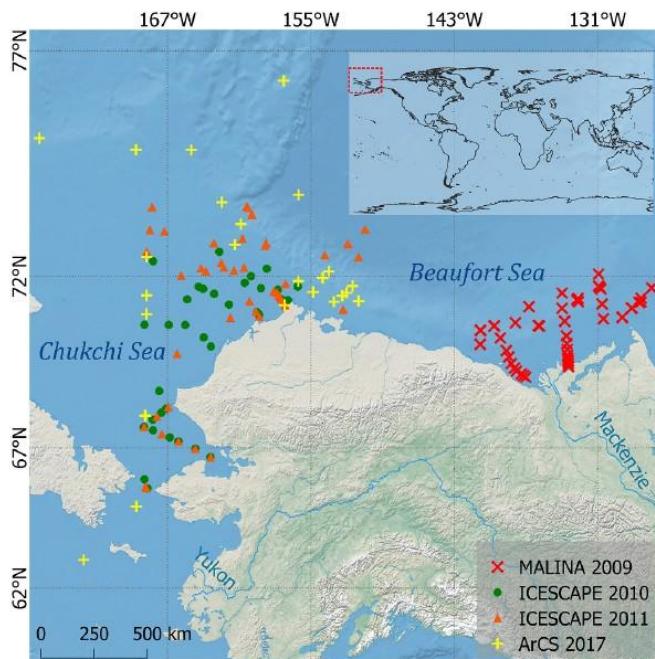
**Etape 3** : Pour développer et valider une relation entre une propriété optique et une variable biogéochimique, il est nécessaire de créer une base de données de *matchups*.



La base de données de *matchups* peut être composée soit de :

- Données optiques ( $R_{rs}$ , IOPs) **in situ** + données biogéochimiques **in situ**
- Données optiques ( $R_{rs}$ , IOPs) **satellite** + données biogéochimiques **in situ**

**Exemple** : base de données de *matchups* entre les mesures *in situ* d'IOPs, de  $R_{rs}(\lambda)$  de POC



Jeu de données dans la région de l'océan Arctique

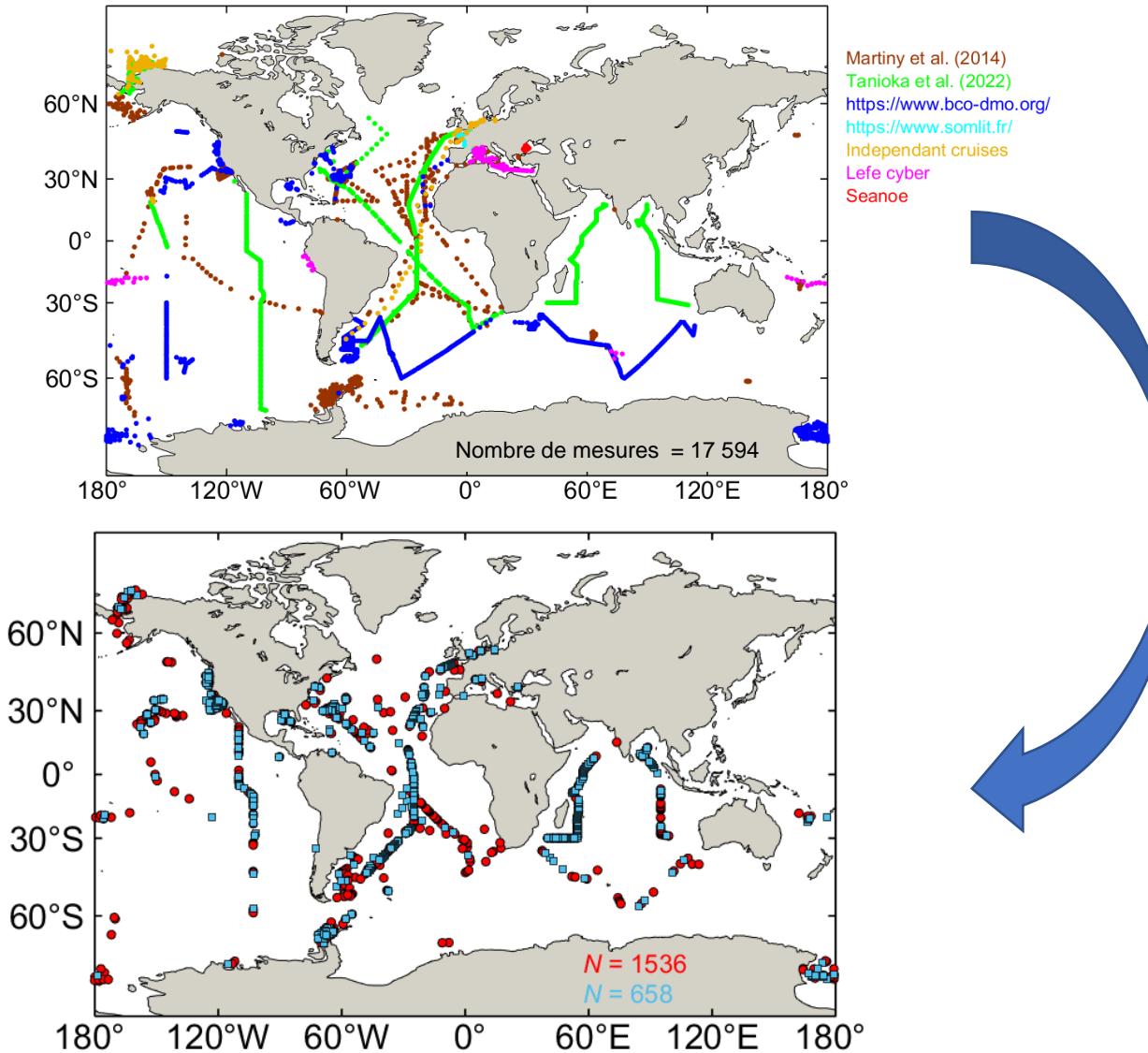
### **Objectif de l'étude:**

L'objectif ici est de tester les effets d'une forte complexité et variabilité de la composition des constituants présents dans l'océan sur le développement d'algorithmes permettant d'estimer le POC.

**Figure 4.** Localisation des mesures concomitantes *in situ* de propriétés optiques inhérentes,  $IOP(\lambda)$ , de reflectance,  $R_{rs}(\lambda)$  et de carbone organique particulaire (POC) dans la région de l'océan Arctique (Stramski et al. 2023).

## 2) Procédure de développement et validation d'un algorithme

**Exemple** : base de données de *matchups* entre les mesures satellite de  $R_{rs}(\lambda)$  et de POC

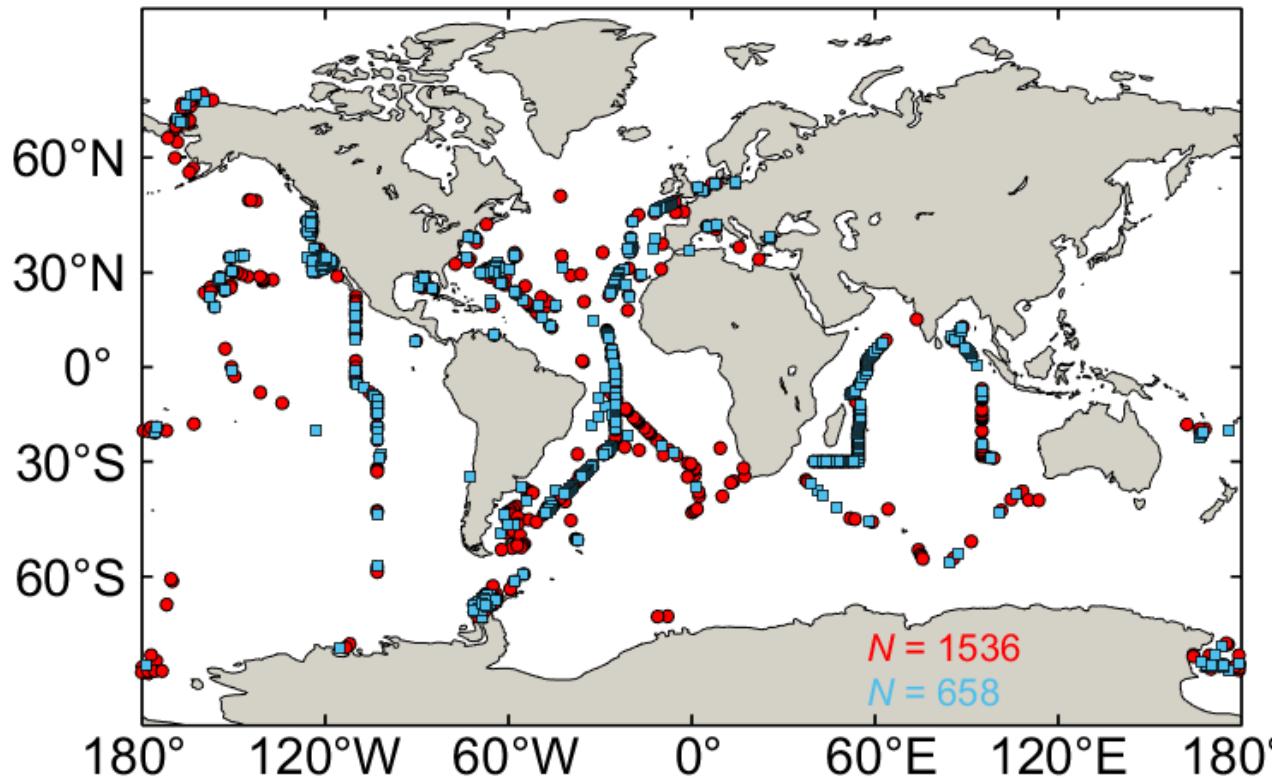


Données optiques provenant de mesures satellites réalisées au « **même endroit** » et en « **même temps** » que les mesures in situ

Figure 5. Base de données de *matchups* entre le POC et les mesures satellites pour le produit Merged.

## 2) Procédure de développement et validation d'un algorithme

### Division de la base de données de *matchups*



**Figure 5.** Base de données de *matchups* entre le POC et les mesures satellites pour le produit Merged. Les marqueurs rouges et bleus représentent respectivement les données utilisées pour le développement et la validation de l'algorithme.

70 % du jeu de données initial va être utilisé pour le développement de l'algorithme

30 % pour la validation



Les 2 jeux de données doivent avoir une distribution similaire

## 2) Procédure de développement et validation d'un algorithme

### Matchups entre les mesures in situ et les données satellites (NASA Ocean Color protocol, Bailey and Wang, 2001)

Données in situ

Trouver le pixel le plus proche de la mesure in situ

Différence temporelle < 3h ?

Exclure les pixels non valides

cv < 15%

La résolution spatiale de la plupart des satellites de mesure de la couleur de l'océan varie entre 300 m et 1.1 km

- Compte tenu de cette échelle, les données in situ sont idéalement collectées dans des régions où la **variabilité spatiale** du paramètre biogéochimique considéré est **relativement stable** sur une zone plusieurs fois supérieure à la résolution spatiale de l'instrument satellitaire.
- Cela permet de minimiser l'effet de la variabilité spatiale à petite échelle sur les données mesurées in situ
- Une zone d'un certain nombre de pixels (par exemple **3x3**) est définie, centrée sur l'emplacement de la mesure in situ.
- Cette boîte permet de générer des statistiques simples, telles que la **moyenne** et **l'écart type**, pour aider à l'évaluation de la stabilité spatiale, ou de l'homogénéité, au point de validation.

X = localisation de la mesure in situ

Latitude	$R_{rs}$	$R_{rs}$	$R_{rs}$
	$R_{rs}$	X	$R_{rs}$
	$R_{rs}$	$R_{rs}$	$R_{rs}$

Longitude

Macro pixel = 9 pixels

## 2) Procédure de développement et validation d'un algorithme

Matchups entre les mesures in situ et les donnée satellites  
(NASA Ocean Color protocol, Bailey and Wang, 2001)

Données in situ

Trouver le pixel le plus proche de la mesure in situ

Différence temporelle  $< 3\text{h}$  ?

Exclure les pixels non valides

cv  $< 15\%$

- Etant donné que les données in situ sont rarement collectées au moment précis où le satellite détecte leur position, nous attribuons un **seuil temporel** à notre définition de la coïncidence.
- **Différence temporelle entre les mesures in situ et satellites  $< 3\text{ h}$**  autour du passage du satellite.
- Les conditions atmosphériques restent raisonnablement stables sur cette période

## 2) Procédure de développement et validation d'un algorithme

Matchups entre les mesures in situ et les données satellites  
(NASA Ocean Color protocol, Bailey and Wang, 2001)

Données in situ

Trouver le pixel le plus proche de la mesure in situ

Différence temporelle < 3h ?

Exclure les pixels non valides

CV < 15%

- Etant donné que les données in situ sont rarement collectées au moment précis où le satellite détecte leur position, nous attribuons un **seuil temporel** à notre définition de la coïncidence.
- **Différence temporelle entre les mesures in situ et satellites < 3 h** autour du passage du satellite.
- Les conditions atmosphériques restent raisonnablement stables sur cette période
- Exclure des valeurs aberrantes (nuage, terre, correction atmosphérique défaillante,  $R_{rs} < 0...$ )
- **5 pixels valides sur les 9**

## 2) Procédure de développement et validation d'un algorithme

Matchups entre les mesures in situ et les données satellites  
(NASA Ocean Color protocol, Bailey and Wang, 2001)

Données in situ

Trouver le pixel le plus proche de la mesure in situ

Différence temporelle < 3h ?

Exclure les pixels non valides

CV < 15%

- Etant donné que les données in situ sont rarement collectées au moment précis où le satellite détecte leur position, nous attribuons un **seuil temporel** à notre définition de la coïncidence.
- **Différence temporelle entre les mesures in situ et satellites < 3 h** autour du passage du satellite.
- Les conditions atmosphériques restent raisonnablement stables sur cette période
- Exclure des valeurs aberrantes (nuage, terre, correction atmosphérique défaillante,  $R_{rs} < 0\dots$ )
- **5 pixels valides sur les 9**
- **Coefficient de variations** ( $CV = (\text{écart-type}/\text{moyenne}) \times 100$ ) des valeurs de  $R_{rs}(\lambda)$  **< 15 %**

## 2) Procédure de développement et validation d'un algorithme

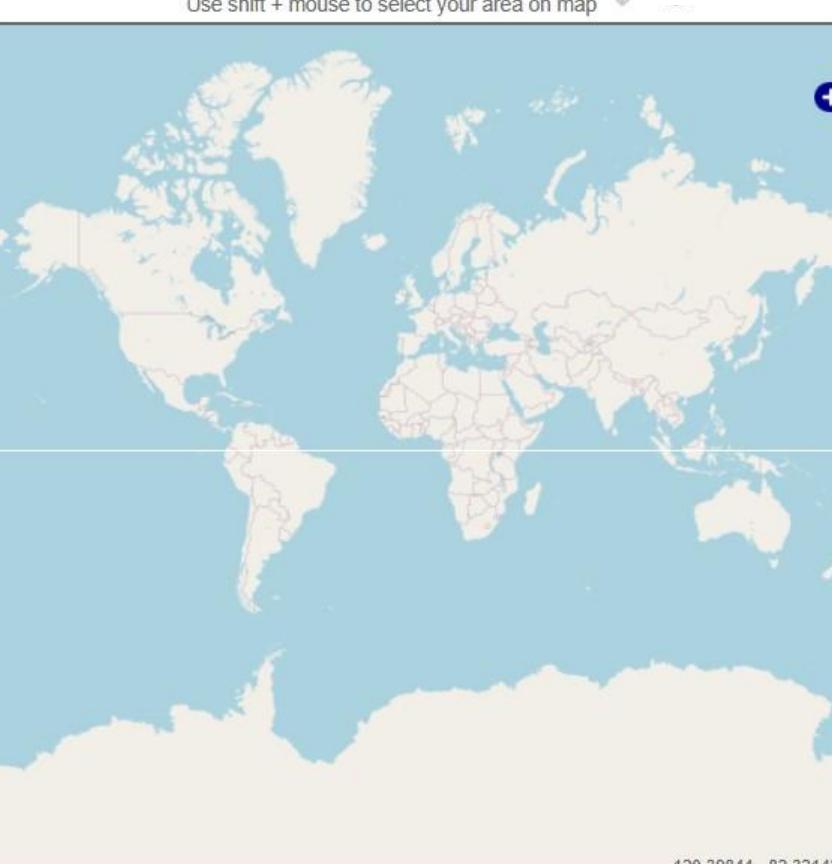
Où récupérer des données satellites (éventuellement pour votre stage) ?

Agency	Data Distribution	URL
NASA	Ocean Color Web Ocean Productivity	<a href="https://oceancolor.gsfc.nasa.gov/">https://oceancolor.gsfc.nasa.gov/</a> <a href="https://www.science.oregonstate.edu/ocean.productivity/">https://www.science.oregonstate.edu/ocean.productivity/</a>
NOAA	CoastWatch	<a href="https://coastwatch.noaa.gov/">https://coastwatch.noaa.gov/</a>
EUMETSAT	EUMETSAT broadcast delivery system (EUMETCast)	<a href="https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETCast/">https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETCast/</a>
EUMETSAT	EUMETSAT Data Centre	<a href="https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETSATDataCentre/index.html">https://www.eumetsat.int/website/home/Data/DataDelivery/EUMETSATDataCentre/index.html</a>
EUMETSAT and ESA	Copernicus Online Data Access (CODA) and SciHub	<a href="https://coda.eumetsat.int">https://coda.eumetsat.int</a> <a href="https://scihub.copernicus.eu/">https://scihub.copernicus.eu/</a>
European Commission	Copernicus Marine Environmental Monitoring Service	<a href="http://marine.copernicus.eu/">http://marine.copernicus.eu/</a>
ESA and PML	Ocean Colour - Climate Change Initiative (OC-CCI)	<a href="https://www.oceancolour.org/">https://www.oceancolour.org/</a>
ACRI-ST	GlobColour	<a href="http://globcolour.info/">http://globcolour.info/</a>

## 2) Procédure de développement et validation d'un algorithme

### Ou récupérer des données satellites (éventuellement pour votre stage) ?

Use shift + mouse to select your area on map



Area

Global  Europe

North: 90  
West: -180 | 180 | East  
South: -90

Projection

Sinusoïdal (L3b)   Plate carrée (L3m) 

Resolution

4 km  25 km  100 km

Date or period

01/09/1997 to 21/08/2025

Binning period

daily  8-days  monthly

Sensor type

merged  meris  modis  seawifs  viirs  
 olcia  olcib  viirsj1

Select product parameters

Check/Uncheck All

Biochemical

CHL1  CHL2  TSM  PIC  POC  
 NFLH  CHL-OC5(\*)  SPM-OC5  PP(\*)

Atmospheric Optical

WVCS  T865  A865  T443  A443  
 T550  A550  CF  ABSD

Ocean Surface Optical

NRRS400  NRRS412  NRRS443  NRRS469  NRRS490  
 NRRS510  NRRS531  NRRS547  NRRS551  NRRS555  
 NRRS560  NRRS620  NRRS645  NRRS670  NRRS674  
 NRRS678  NRRS681  RRS681  NRRS709  RRS709  
 NRRS711  NRRS712  NRRS713  NRRS714  NRRS715

Choix de la région

Résolution spatiale

Date

Résolution temporelle

Capteurs

Variables biogéochimiques  
Chla, POC, SPM...

$R_{rs}(\lambda)$



[https://hermes.acri.fr/images/GlobCOLOUR\\_PUG.pdf](https://hermes.acri.fr/images/GlobCOLOUR_PUG.pdf)  
<https://hermes.acri.fr/index.php?class=archive>

## 2) Procédure de développement et validation d'un algorithme

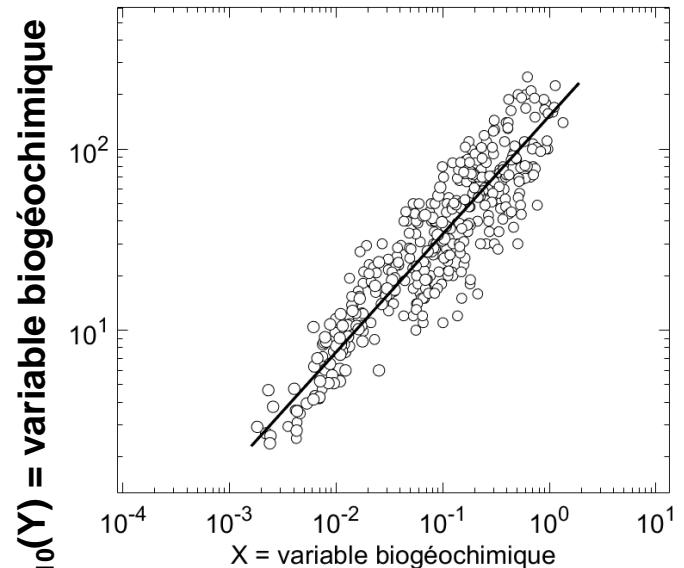
### Etape 4 : Choix du type de modèle de régression



Le **Modèle-I de régression linéaire** (Laws, 1997)

- Le plus fréquemment utilisé
- Basé sur l'hypothèse que **seulement la variable Y a une incertitude**
- La variable X est supposée ne pas avoir d'erreur associée

**L'équation nous indique comment Y varie en réponse aux changements de X.**



**log<sub>10</sub>(X) = variable optique**

**Figure 6.** Relation entre une variable biogéochimique et une variable optique.

**log<sub>10</sub>(Y) = Slope × log<sub>10</sub>(X) + Intercept**

Ce qui correspond à une fonction puissance de type :

**Y = A × X<sup>B</sup>**

A et B = coefficients de régression d'une fonction puissance

**A = 10<sup>Intercept</sup>**

**B = Slope**

## 2) Procédure de développement et validation d'un algorithme

### Etape 4 : Choix du type de modèle de régression



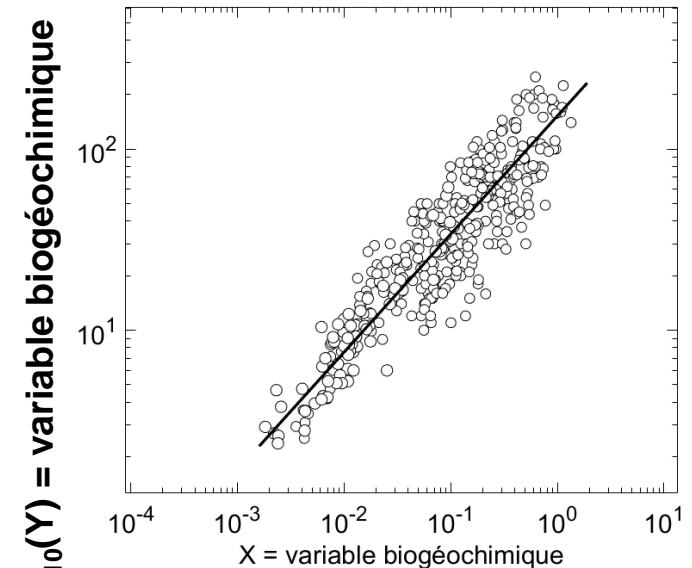
#### Le Modèle-I de régression linéaire (Laws, 1997)

- Le plus fréquemment utilisé
- Basé sur l'hypothèse que **seulement la variable Y a une incertitude**
- La variable X est supposée ne pas avoir d'erreur associée

**L'équation nous indique comment Y varie en réponse aux changements de X.**

#### Le Modèle-II de régression linéaire (Press et al., 1992; Laws, 1997)

- Quantifie la force d'une relation linéaire entre deux variables. Cette méthode assume que les 2 variables, X et Y, sont associées à des incertitudes,
- Sous certaines conditions, cette méthode peut être utilisée dans un but prédictif, notamment lorsque l'erreur associée à X n'est pas significativement inférieure à l'erreur associée à Y (McArdle, 1988).



**$\log_{10}(Y) = \text{Slope} \times \log_{10}(X) + \text{Intercept}$**

**Figure 6.** Relation entre une variable biogéochimique et une variable optique.

**$\log_{10}(Y) = \text{Slope} \times \log_{10}(X) + \text{Intercept}$**

Ce qui correspond à une fonction puissance de type :

**$Y = A \times X^B$**

$A$  et  $B$  = coefficients de régression d'une fonction puissance

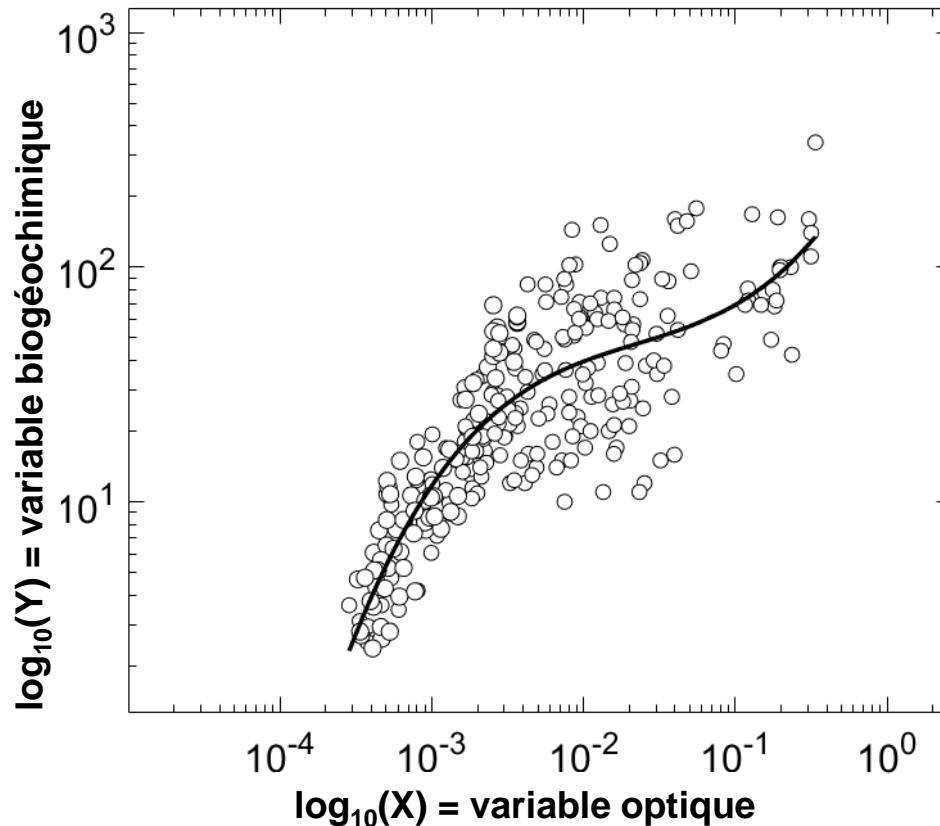
**$A = 10^{\text{Intercept}}$**

**$B = \text{Slope}$**

## 2) Procédure de développement et validation d'un algorithme

Dans certains cas, le modèle général ne peut pas être représenté par une fonction puissance.

→ Fonction polynomiale de *i* degrés



**Figure 7.** Relation entre une variable biogéochimique et une variable optique. La ligne noire continue représente la fonction polynomiale du troisième degré la mieux ajustée pour l'ensemble des données.

$$\log_{10}(Y) = a_0 + \sum_{i=1}^N a_i \log_{10}(X)^i$$
$$Y = 10^{[a_0 + \sum_{i=1}^N a_i \times \log_{10}(X)^i]}$$

**Exemple** : relation polynomiale de 3<sup>em</sup>-degrés

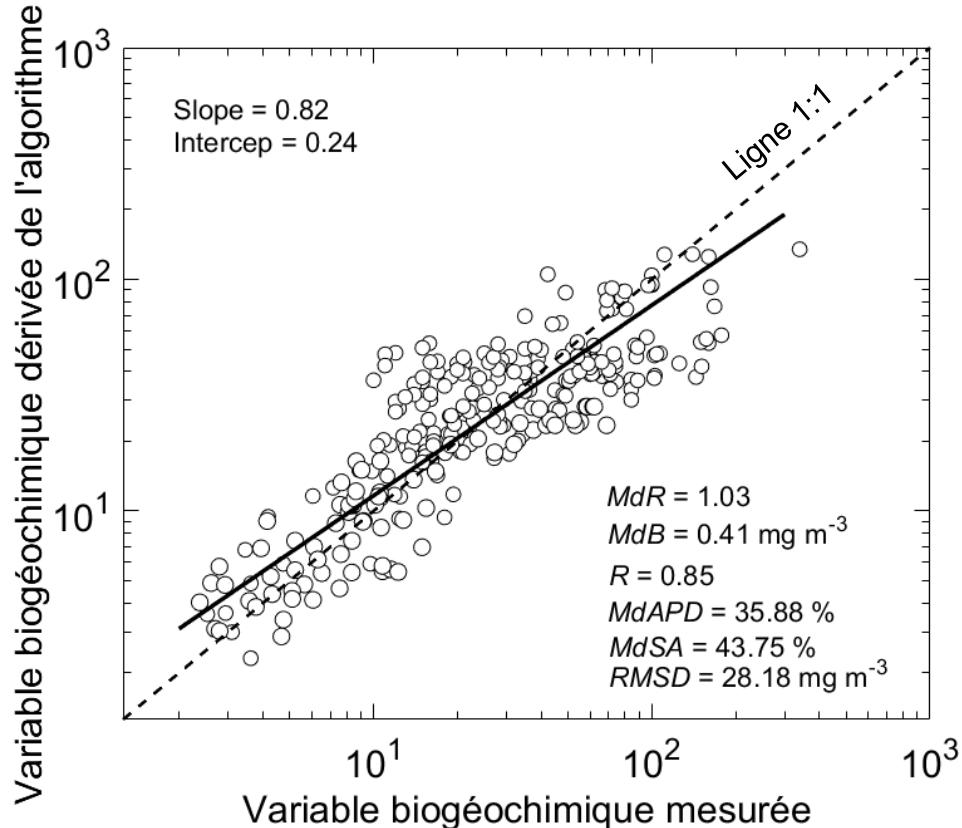
$$Y = 10^{[a_0 + a_1 [\log_{10}(X)] + a_2 [\log_{10}(X)]^2 + a_3 [\log_{10}(X)]^3]}$$

$a_0, a_1, a_2$ , et  $a_3$  = coefficients de régression

## 2) Procédure de développement et validation d'un algorithme

### Etape 5 : Test de performance et validation de l'algorithme

Pour **évaluer**, **valider**, ou **comparer** la performance d'un algorithme, on applique l'algorithme à des mesures optiques indépendantes, puis on les compare à des valeurs mesurées in situ, en utilisant un **modèle de régression II**.



On utilise des **paramètres statistiques** permettant de caractériser les **différences systématiques et relatives** entre les valeurs dérivées d'un algorithme et les valeurs mesurées.

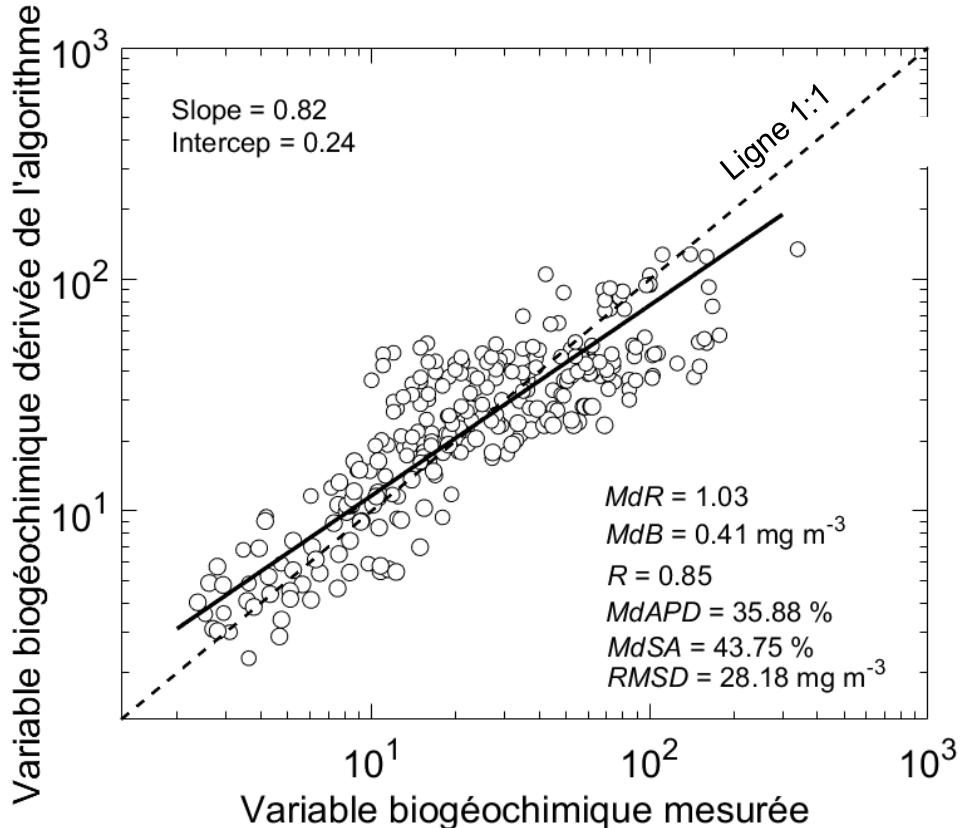
**Question** : Sur cette figure, que vous indique la valeur de la pente ?

**Figure 8.** Relation entre la variable biogéochimique dérivée de l'algorithme et mesurée.

## 2) Procédure de développement et validation d'un algorithme

### Etape 5 : Test de performance et validation de l'algorithme

Pour **évaluer**, **valider**, ou **comparer** la performance d'un algorithme, on applique l'algorithme à des mesures optiques indépendantes, puis on les compare à des valeurs mesurées in situ, en utilisant un **modèle de régression II**.



**Figure 8.** Relation entre la variable biogéochimique dérivée de l'algorithme et mesurée.

On utilise des **paramètres statistiques** permettant de caractériser les **différences systématiques et relatives** entre les valeurs dérivées d'un algorithme et les valeurs mesurées.

**Question** : Sur cette figure, que vous indique la valeur de la pente ?

- **Sur estimation des faible valeurs** dérivées de l'algorithme par rapport aux valeurs mesurées
- **Sous estimation des fortes valeurs** dérivées de l'algorithme par rapport aux valeurs mesurées

## 2) Procédure de développement et validation d'un algorithme

### Etape 5 : Test de performance et validation de l'algorithme

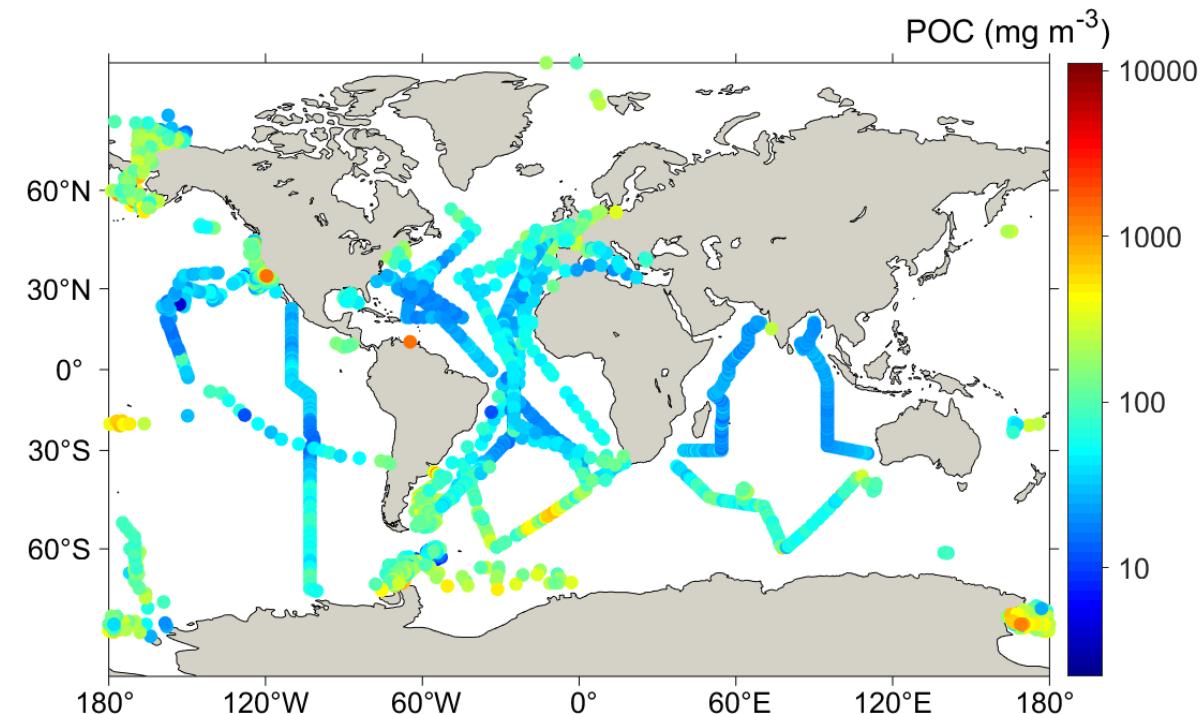
Pour comparer et évaluer la performance d'un algorithme, on utilise des **paramètres statistiques** permettant de caractériser les **différences systématiques et relatives** entre les valeurs dérivées d'un algorithme et les valeurs mesurées.

Symbol	Description	
$y_i, x_i$	Algorithm-derived variable ( $y_i$ ) and measured variable ( $x_i$ ) for sample $i$ of $N$	
$N$	Number of samples (data)	
$R$	Pearson's product moment correlation coefficient between $\log_{10}$ -transformed variables used in model-II linear regression	
$S$ and $I$	Slope and intercept obtained from model linear regression	Reveal the potential presence of bias across the dynamic range of variable.
$MdB$ (mg m <sup>-3</sup> )	Median Bias; median value of $(y_i - x_i)$	Measure of aggregate bias
$MdR$	Median Ratio of $(y_i/x_i)$	Measure of aggregate bias
$MdAPD$ (%)	Median Absolute Percentage Difference; median value of $100 \times [  (y_i - x_i) / x_i  ]$	Measure of relative difference
$MdSA$ (%)	Median Symmetric Accuracy; $100 \times [10^{\text{median}[\log(y_i/x_i)]} - 1]$	Measure of relative difference
$RMSD$ (mg m <sup>-3</sup> )	Root Mean Square Deviation; $[(1/N) \sum^N (y_i - x_i)^2]^{0.5}$	Characterize random deviations

### 3) Cas d'étude : le carbone organique particulaire

#### 3.1) Le POC océanique

Connaitre les variations temporelles et spatiales de POC dans l'océan permet de mieux **estimer la biomasse** de carbone du phytoplancton, la **production primaire**, et le **budget total** de carbone dans l'océan, ainsi que de mieux comprendre **l'export de carbone** de l'océan de surface vers le fond (Allison et al., 2010).



**Figure 9.** Localisation géographique des mesures de carbone organique particulaire dans l'océan.

Les **données in situ** représentent ce qui se rapproche le plus de la « **réalité** » à un moment donné, cependant la **couverture spatiale et temporelle est faible** en comparaison avec les estimations de POC à partir de données satellites.

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

Le premier algorithme permettant d'estimer le POC a été développé en 1999 à partir de mesures in situ de réflectance  $R_{rs}(\lambda)$ , du coefficient total de rétrodiffusion  $b_b(\lambda)$ , réalisées dans la région Antarctique.

##### Approche basée sur deux algorithmes empiriques

La première étape repose sur la dépendance du  $b_{bp}(510)$  avec les particules en suspension

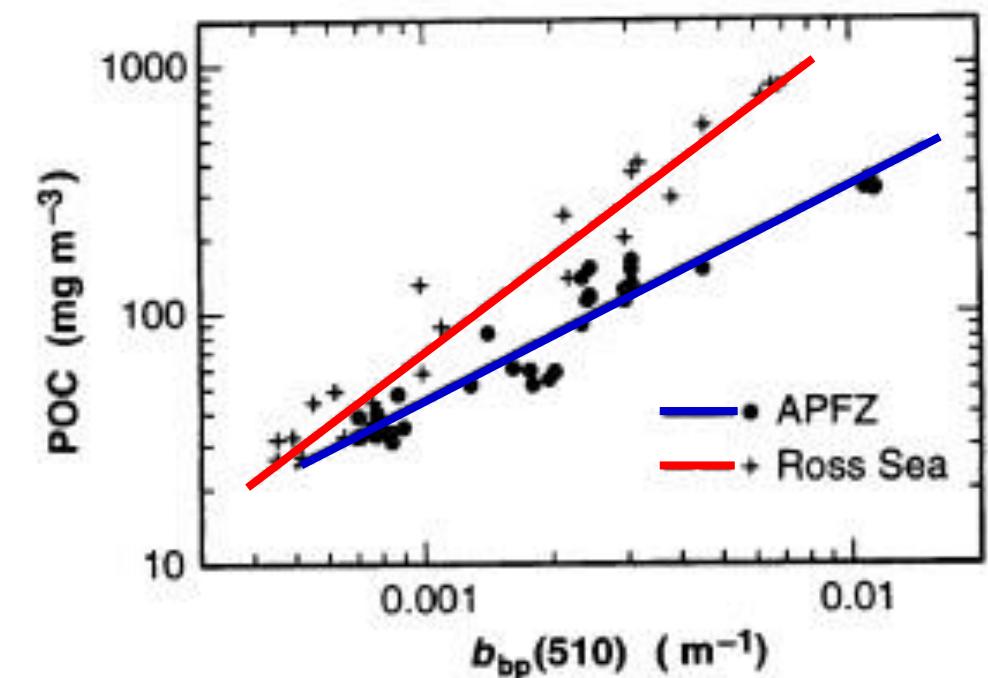
- Forte corrélation entre le  $b_{bp}$  et le POC

Zone du front polaire (APFZ)

$$\text{POC} = 17069 \cdot b_{bp}(510)^{0.859}$$

$R^2 = 0.92, \quad N = 33$

- La forte corrélation est due à la prédominance de la concentration en particules organiques dans le contrôle des variations à la fois du POC et du  $b_{bp}$ .
- Les relations diffèrents entre la région APFZ et la mer de Ross.

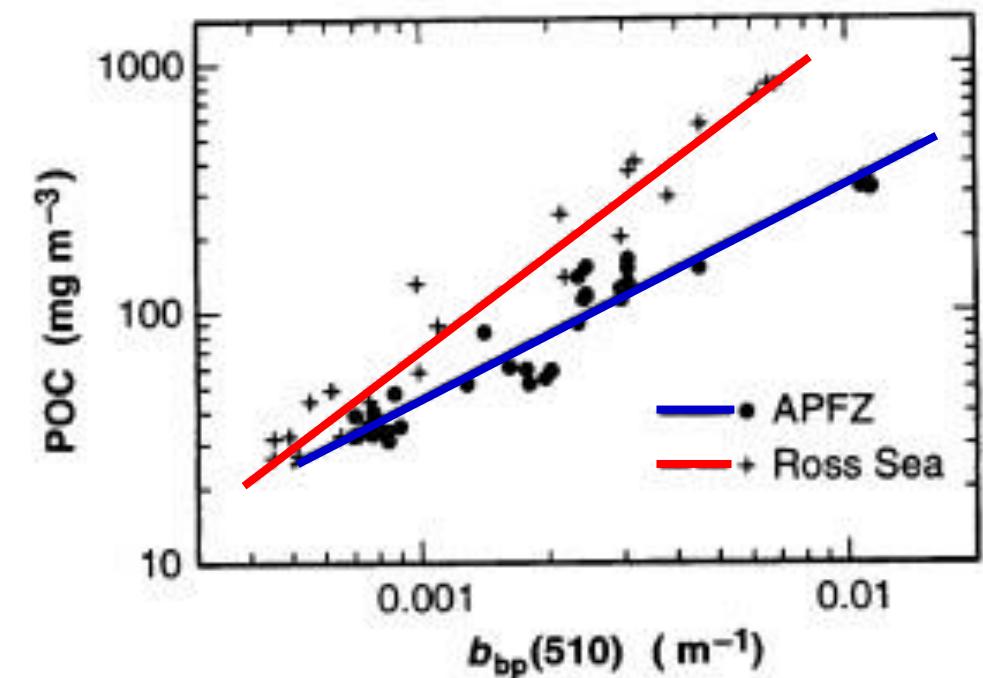


**Figure 10.** Relation entre le POC et le  $b_{bp}(510)$ .  
Figure tirée de Stramski et al., (1999).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

**Question :** Comment expliquer le fait que les relations diffèrent entre la mer de Ross et la région de l'APFZ ?



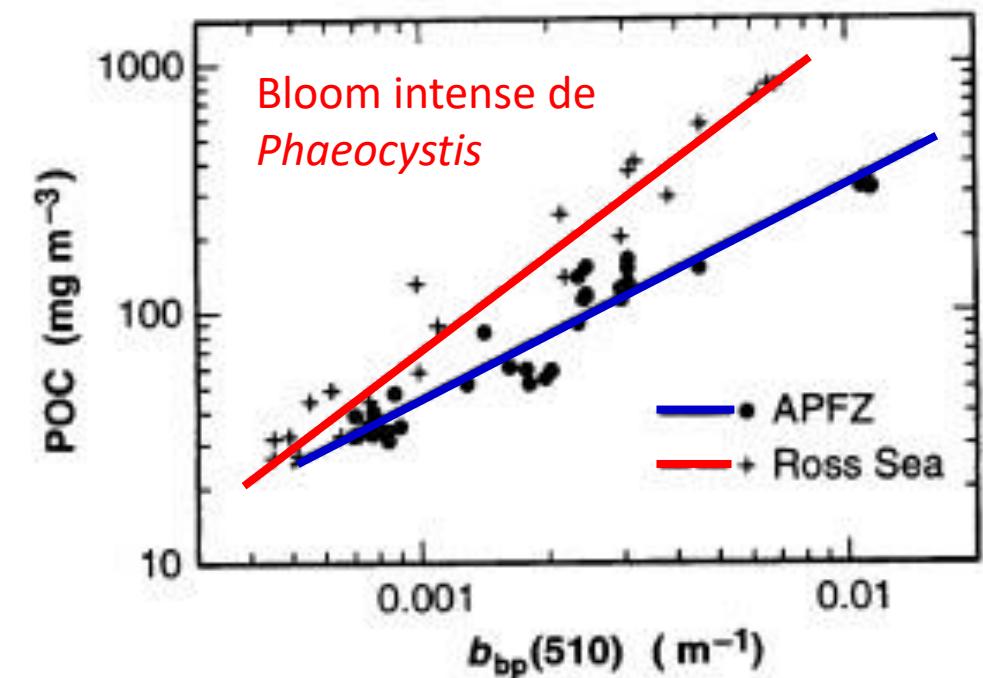
**Figure 10.** Relation entre le POC et le  $b_{bp}(510)$ .  
Figure tirée de Stramski et al., (1999).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

**Question :** Comment expliquer le fait que les relations diffèrent entre la mer de Ross et la région de l'APFZ ?

- La composition de la matière particulaire varie dans le temps et d'une région à l'autre, ce qui peut expliquer que l'on observe des relations différentes entre les deux régions.
- En mer de Ross, les mesures ont été réalisées lors d'un bloom de *Phaeocystis*.
- Type de phytoplancton qui rétrodiffuse moins que les autres particules.



**Figure 10.** Relation entre le POC et le  $b_{bp}(510)$ .  
Figure tirée de Stramski et al., (1999).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

La première étape repose sur la dépendance du  $b_{bp}(510)$  avec les particules en suspension

$$\text{POC} = 17069 \cdot b_{bp}(510)^{0.859}$$

$$R^2 = 0.92, \quad N = 33$$

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

La première étape repose sur la dépendance du  $b_{bp}(510)$  avec les particules en suspension

$$\text{POC} = 17069 \cdot b_{bp}(510)^{0.859} \quad R^2 = 0.92, \quad N = 33$$

La seconde étape repose sur la dépendance des valeur de  $R_{rs}(555)$  avec le coefficient total de rétrodiffusion,  $b_b(510)$

$$b_b(\lambda) = b_{bbp}(\lambda) + b_{bw}(\lambda)$$

- Forte corrélation entre le  $b_b(510)$  et la  $R_{rs}(555)$ .
- Les relations entre la région APFZ et la mer de Ross sont proches.

$$b_b(510) = 1756 \cdot R_{rs}(555) - 4.772 \cdot 10^{-4} \quad R^2 = 0.97, \quad N = 14$$

- Sachant que l'on connaît la valeur de  $b_{bw}$  ( $= 0.0013 \text{ m}^{-1}$ ) et en combinant la relation POC vs.  $b_{bp}(510)$  avec la relation  $b_b(510)$  vs.  $R_{rs}(555)$ , on obtient l'algorithme final pour estimer le POC à partir de la  $R_{rs}(555)$ .

$$\text{POC} = 17069 \cdot [1756 \cdot R_{rs}(555) - 1.777 \cdot 10^{-3}]^{0.859}$$

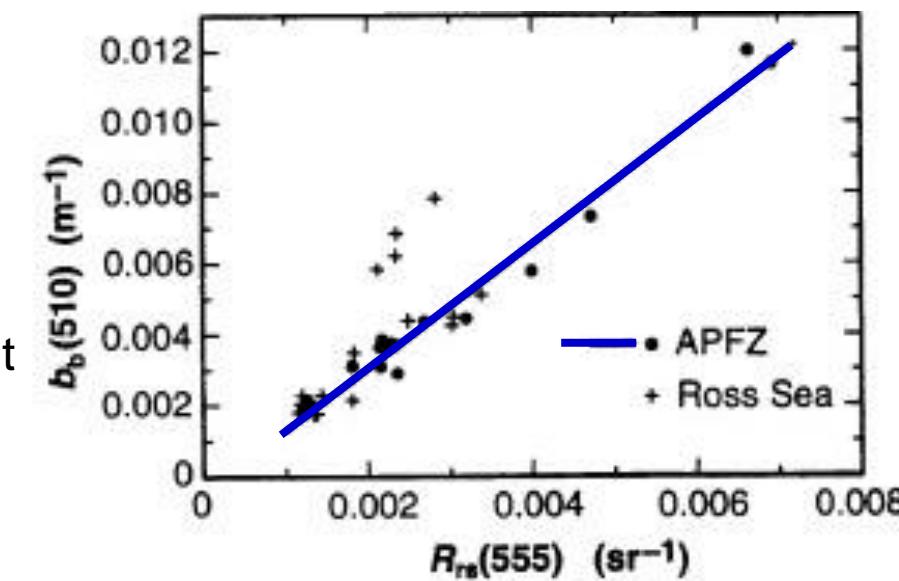


Figure 11. Relation entre le  $b_b(510)$  et la  $R_{rs}(555)$ .  
Figure tirée de Stramski et al., (1999).

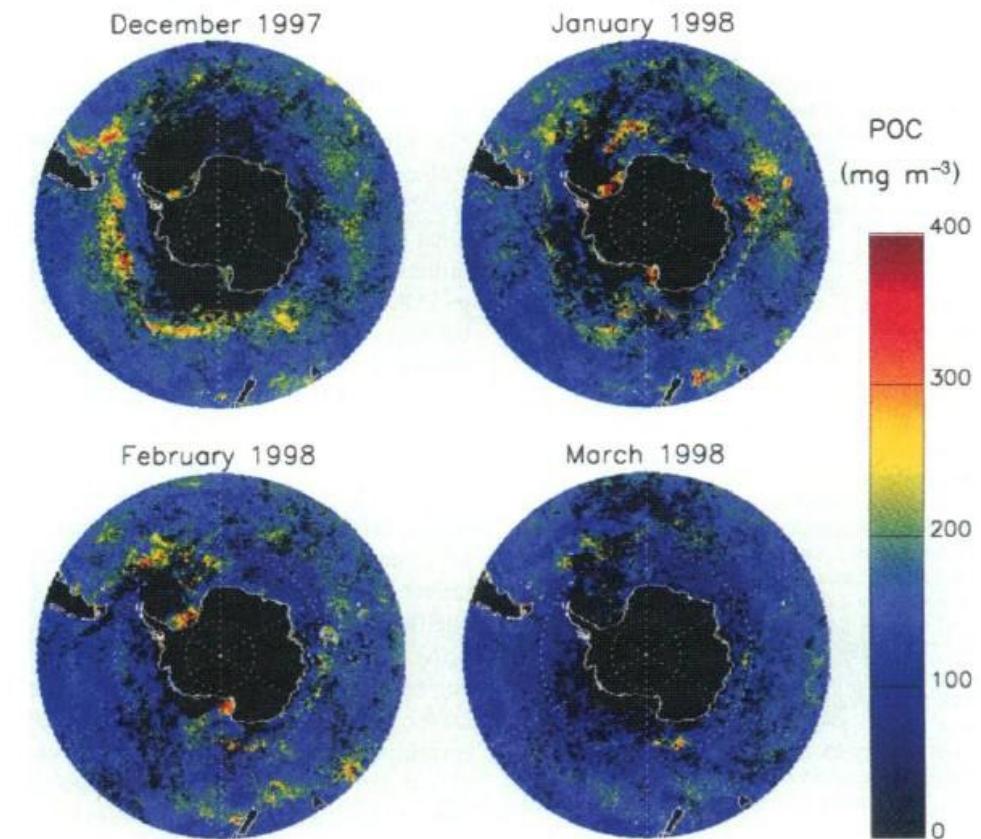
### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

Application de cet algorithme aux données de  $R_{rs}(555)$  satellite du capteur SeaWIFS

Première estimation des variations saisonnières du POC dans les eaux de surface de l'océan Antarctique

Question : Peut-on utiliser cet algorithme pour estimer le POC dans l'océan ouvert ?



**Figure 12.** Estimation du POC à partir de la  $R_{rs}(555)$ .  
Figure tirée de Stramski et al., (1999).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.2) Première estimation du POC (Stramski et al., 1999)

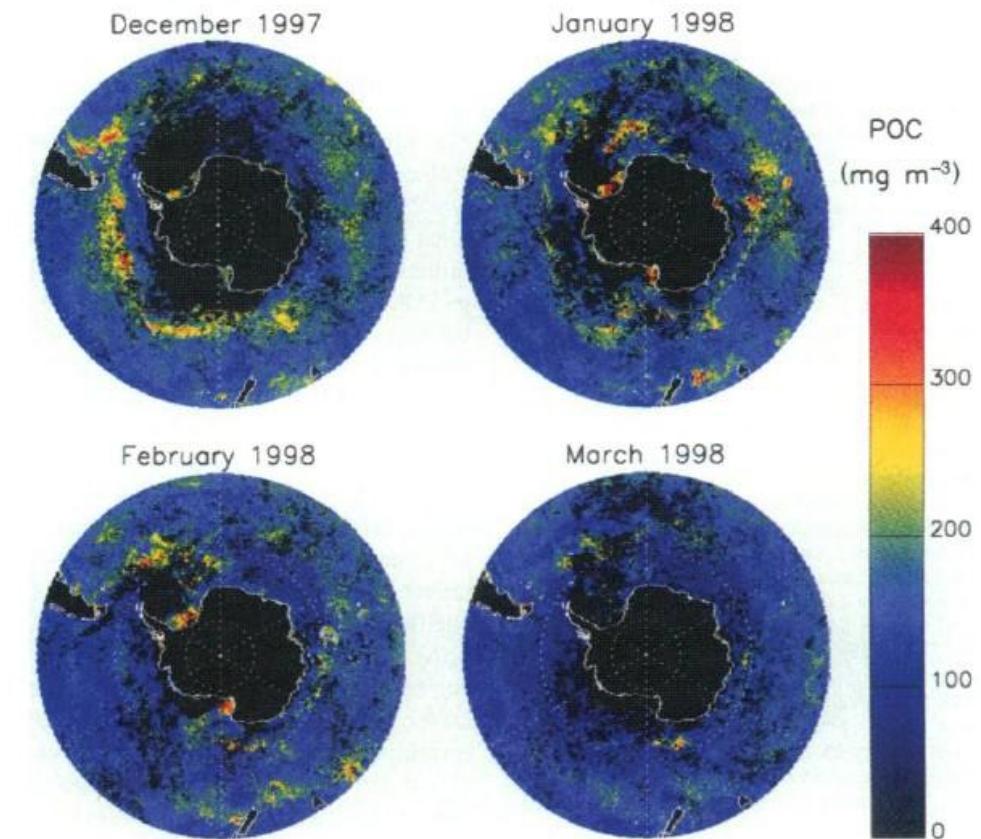
Application de cet algorithme aux données de  $R_{rs}(555)$  satellite du capteur SeaWIFS

Première estimation des variations saisonnières du POC dans les eaux de surface de l'océan Antarctique

Question : Peut-on utiliser cet algorithme pour estimer le POC dans l'océan ouvert ?



- La variété des caractéristiques physiques/chimiques des assemblages des différentes particules a un effet important sur les propriétés optiques.
- Par conséquent, ce jeu de données n'est pas adapté à l'estimation du POC dans les régions contrastées telles que les zones oligotrophes
- Nécessité de développer de nouveaux algorithmes adaptés à l'océan ouvert



**Figure 12.** Estimation du POC à partir de la  $R_{rs}(555)$ .  
Figure tirée de Stramski et al., (1999).

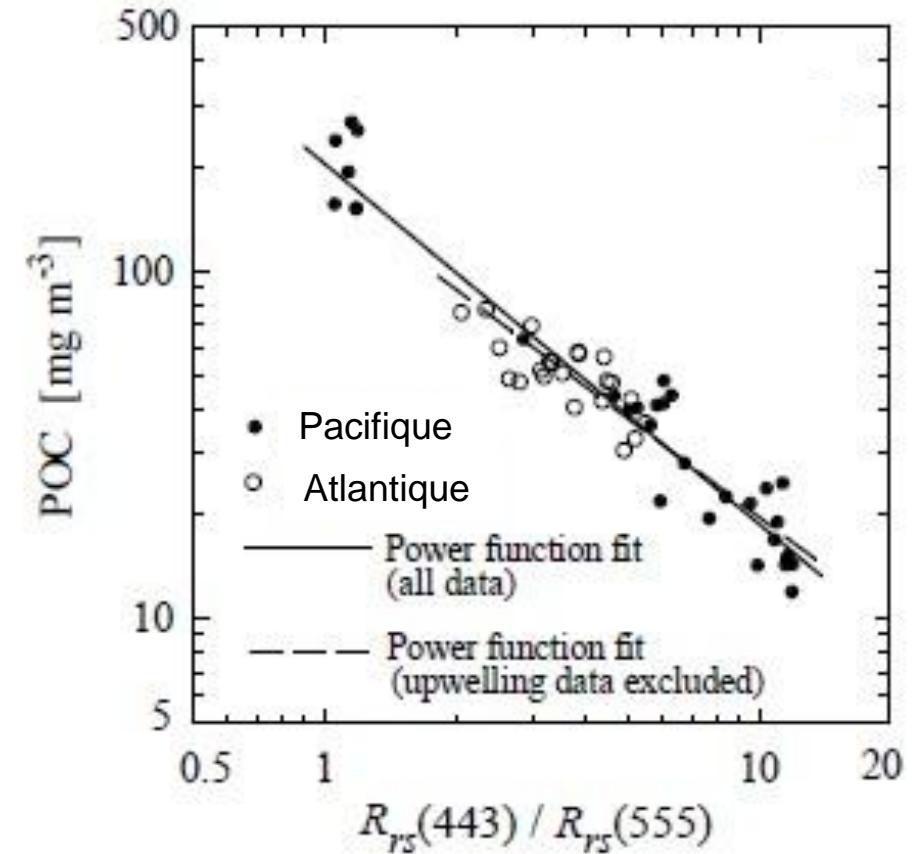
### 3) Cas d'étude : le carbone organique particulaire

#### 3.3) Estimation du POC dans l'océan ouvert

Actuellement, le POC est estimé à partir de l'algorithme empirique développé par Stramski et al. (2008) et implémenté par la NASA (<https://oceancolor.gsfc.nasa.gov/resources/atbd/poc>)

- Cet algorithme est développé à partir d'une base de données de *matchups* de mesures in situ réalisées dans l'océan Pacifique sud et l'océan Atlantique.
- Différentes relations entre le POC et les IOPs, et le POC et la  $R_{rs}$  ont été testées.
- Cet algorithme est basé sur la relation entre le POC et le rapport de  $R_{rs}(443)/R_{rs}(555)$ .
- Relation linéaire sous forme de fonction puissance

$$POC = 203.2 \times \frac{R_{rs}(443)}{R_{rs}(555)}^{-1.034}$$



**Figure 13.** Relation entre le POC et le rapport  $R_{rs}(442)/R_{rs}(555)$ . Figure tirée de Stramski et al., (2008).

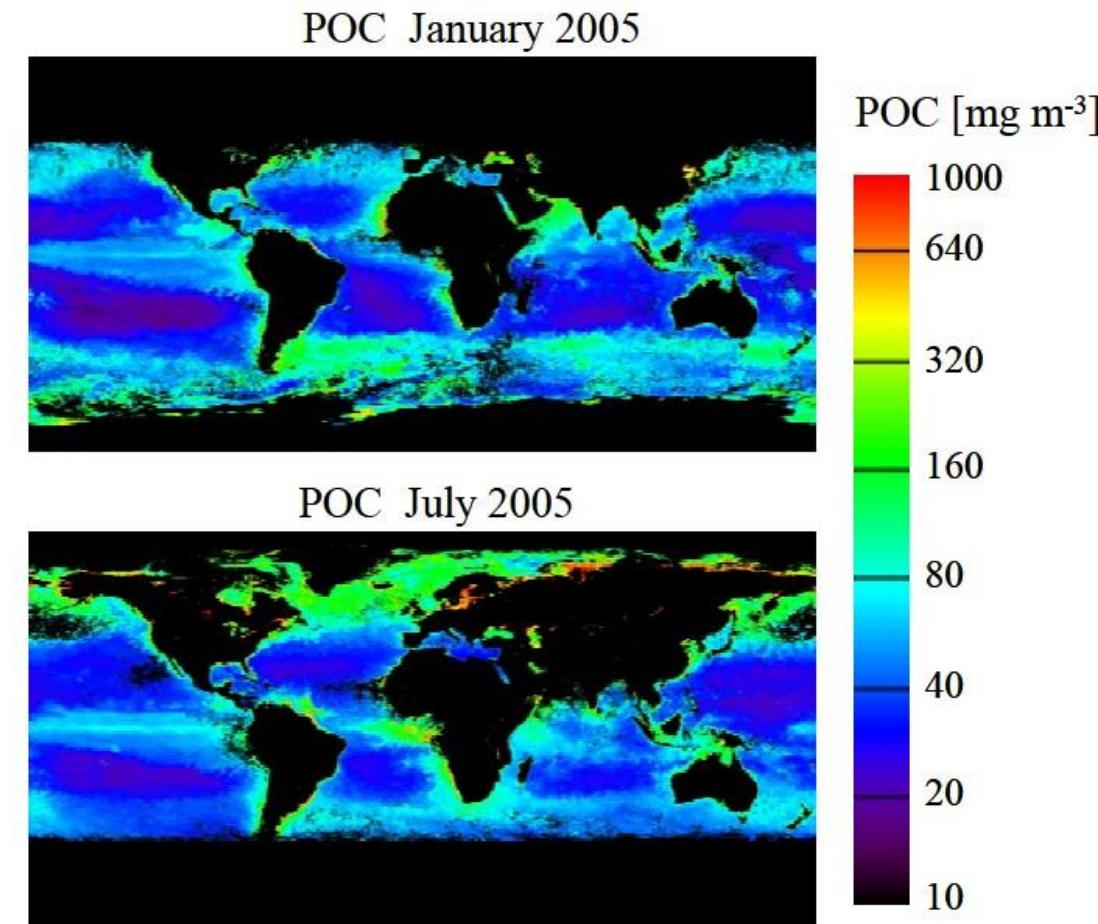
### 3) Cas d'étude : le carbone organique particulaire

#### 3.3) Estimation du POC dans l'océan ouvert

- Valide uniquement dans les eaux de l'océan ouvert (cas 1), (ultra-oligotrophe, oligotrophe, régime d'upwelling côtier)
- Valide sur une large gamme d'assemblage particulaire
- Non valide dans les zones côtières
- Valide pour des C° en POC comprises entre 10 – 270 mg m<sup>-3</sup>
- Couverture temporelle : 1997-aujourd'hui

#### Applications possibles

- Analyse de tendance temporelle
- Budget de POC dans l'océan ouvert
- Informations sur la production primaire



**Figure 14.** Concentrations moyennes mensuelles en POC dérivées des observations satellites avec SeaWiFS.  
Figure tirée de Stramski et al., 2008.

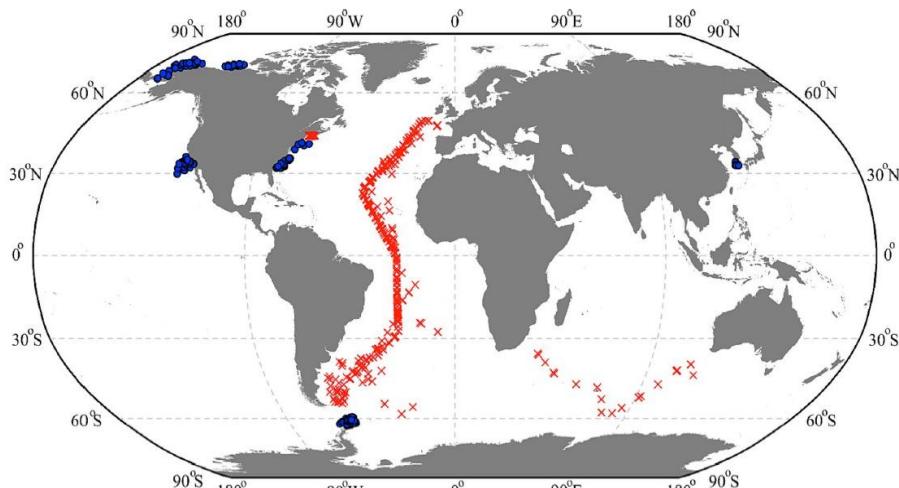
### 3) Cas d'étude : Algorithmes empiriques permettant d'estimer le POC

#### 3.4) Validation du POC dans l'océan ouvert

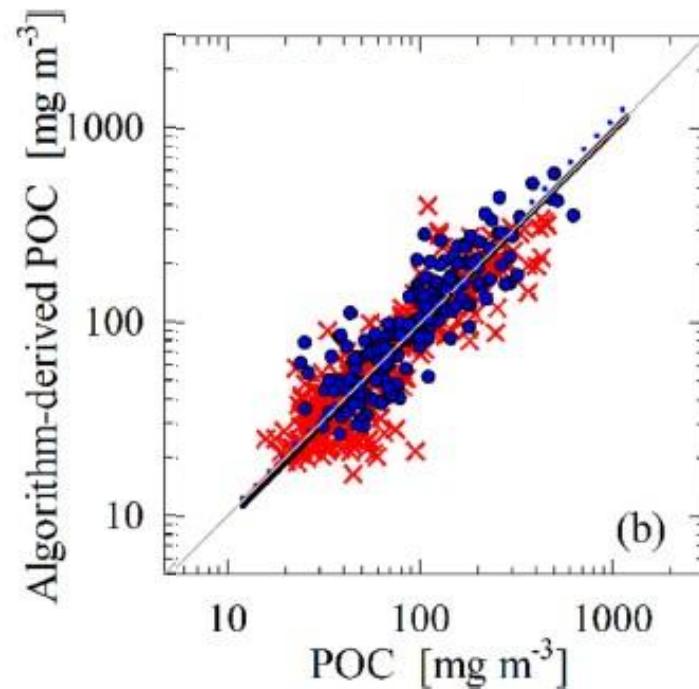


Pour valider un algorithme, on utilise des **paramètres statistiques** permettant de caractériser les **différences systématiques et relatives** entre les valeurs dérivées d'un algorithme et les valeurs mesurées.

Ces paramètres statistiques permettent également de comparer la performance des différents algorithmes existants pour une même variable



**Figure 15.** Localisation géographique du jeu de données de validation. Figure tirée de Stramski et al., 2022.



**Figure 16.** Relation entre le POC dérivé de l'algorithme et le POC mesuré. Figure tirée de Stramski et al., 2022.

#### Paramètres statistiques

Pente ( $S$ ) = 1.01  
Intercept ( $I$ ) = -0.04  
Coefficient de corrélation ( $R$ ) = 0.91  
 $MdAPD$  = 20.1 %  
 $MdB$  = -1.1  $mg\ m^{-3}$   
 $MdR$  = 0.98  
 $MdSA$  = 23.0 %

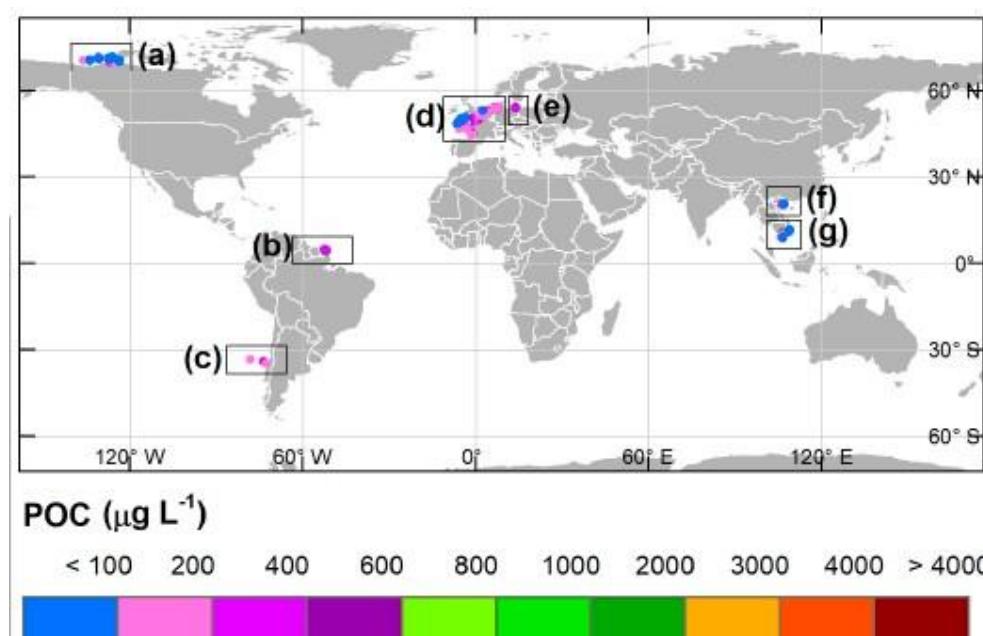
### 3) Cas d'étude : le carbone organique particulaire

#### 3.5) Estimation du POC dans les régions côtières

Au cours des vingt dernières années, de nombreuses études ont permis d'améliorer les algorithmes notamment dans les régions côtières, qui sont beaucoup plus complexes d'un point de vue optique que les régions de l'océan ouvert.

Développement d'algorithme permettant d'estimer le POC dans les régions côtières (Kien et al., 2019)

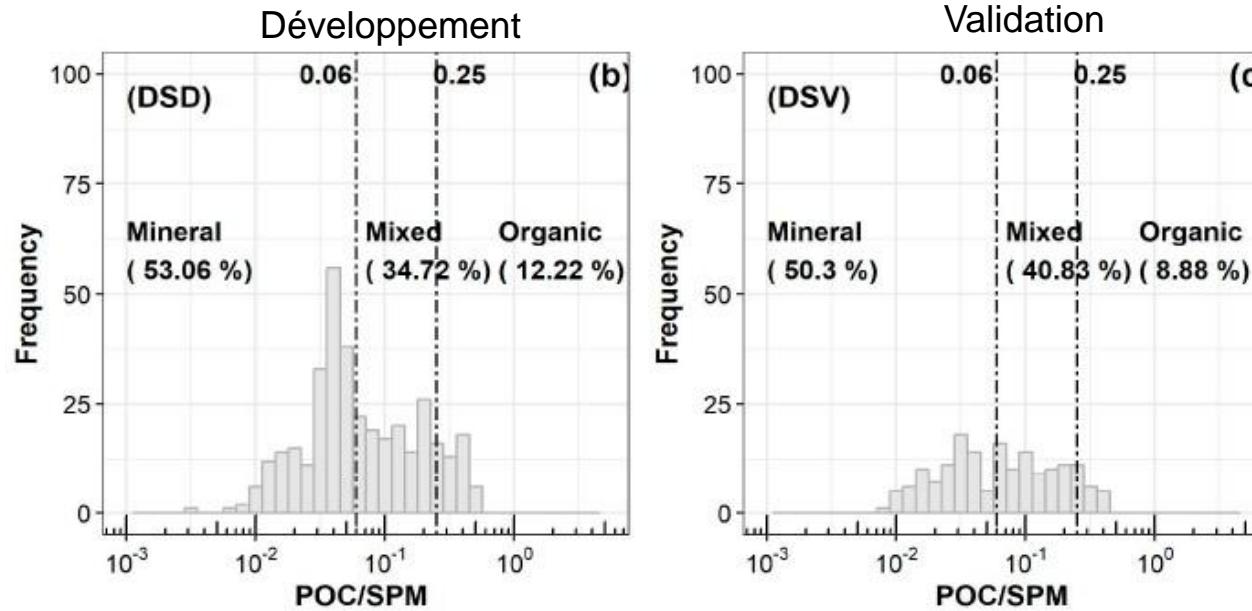
- Base de données de matchups  $R_{rs}/POC$  uniquement dans les régions côtières
- POC [45 – 5744 mg m<sup>-3</sup>]
- N = 606 matchups



**Figure 17.** Localisation des matchups  $R_{rs}/POC$  mesurés dans différentes régions côtières (issue de Kien et al., 2019).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.5) Estimation du POC dans les régions côtières



SPM = matière en suspension  
particulaire = minéral + biomasse



Le rapport POC/SPM peut être utilisé comme un proxy pour caractériser la contribution des particules organiques vs. inorganiques de la matière en suspension.

Figure 18. Nombre d'occurrences des mesures du rapport POC/SPM.

Figure tirée de Kien et al. (2019).

#### Caractérisation du jeu de données

- $\text{POC/SPM} < 0.06 \rightarrow$  dominance du minéral par rapport à la biomasse
- $0.06 < \text{POC/SPM} < 0.25 \rightarrow$  mélange minéral/biomasse
- $\text{POC/SPM} > 0.25 \rightarrow$  dominance biomasse



Ces seuils peuvent être variables d'une étude à l'autre.

### 3) Cas d'étude : le carbone organique particulaire

#### 3.5) Estimation du POC dans les régions côtières

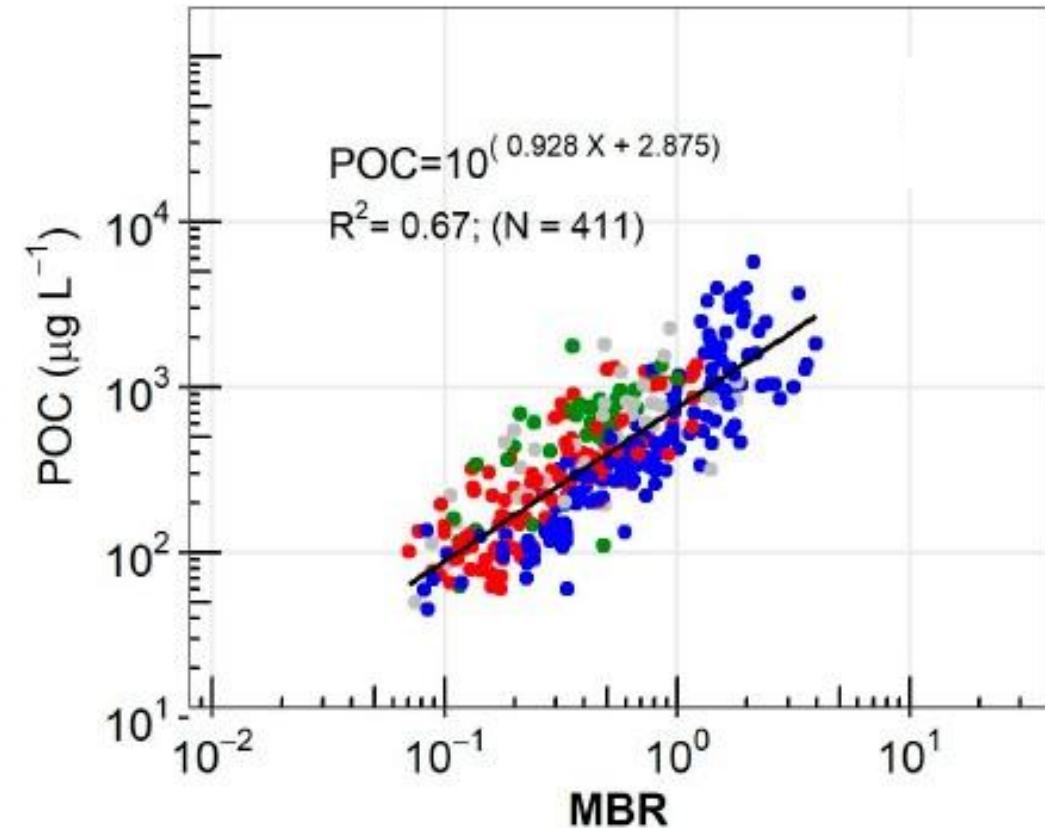
Algorithme empirique basé sur une relation linéaire entre le POC et le MBR (Maximum Band Ratio)

Adaptation des algorithmes développés dans l'océan ouvert, basé également sur les rapports de bandes de  $R_{rs}$

$$POC = 10^{0.928X + 2.875}$$

$$X = \log_{10} \left[ \text{maximum} \left( \frac{R_{rs}(665)}{R_{rs}(490)}, \frac{R_{rs}(665)}{R_{rs}(510)}, \frac{R_{rs}(665)}{R_{rs}(555)} \right) \right]$$

L'utilisation du MBR plutôt que d'un seul rapport de bande permet de maximiser la précision du modèle sur la gamme de variation du POC



**Figure 19.** Relation entre le POC et le MBR.  
Figure tirée de Kien et al. (2019).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif



La forte complexité et la variabilité de la composition des constituants représentent un défi majeur pour le développement d'algorithmes visant à estimer les variables biogéochimiques à partir de mesures optiques couvrant une large gamme d'environnements marins.

Les particules minérales et organiques (notamment les cellules planctoniques) présentent généralement des indices de réfraction très différents (Aas, 1996). De telles différences d'indice de réfraction ont des implications majeures sur l'interaction de la lumière avec les particules (e.g. Babin et al., 2003; Stramski and Kiefer, 1991).

Il est par conséquent très difficile de développer un algorithme bio-optique ou de télédétection unique, quelle que soit son approche, empirique ou semi-analytique, capable d'estimer les concentrations en POC sur un large continuum d'environnements aquatiques.

Question : Quelle serait la possible voie d'amélioration permettant de développer un seul algorithme bio-optique valide sur un continuum de divers environnements aquatiques?

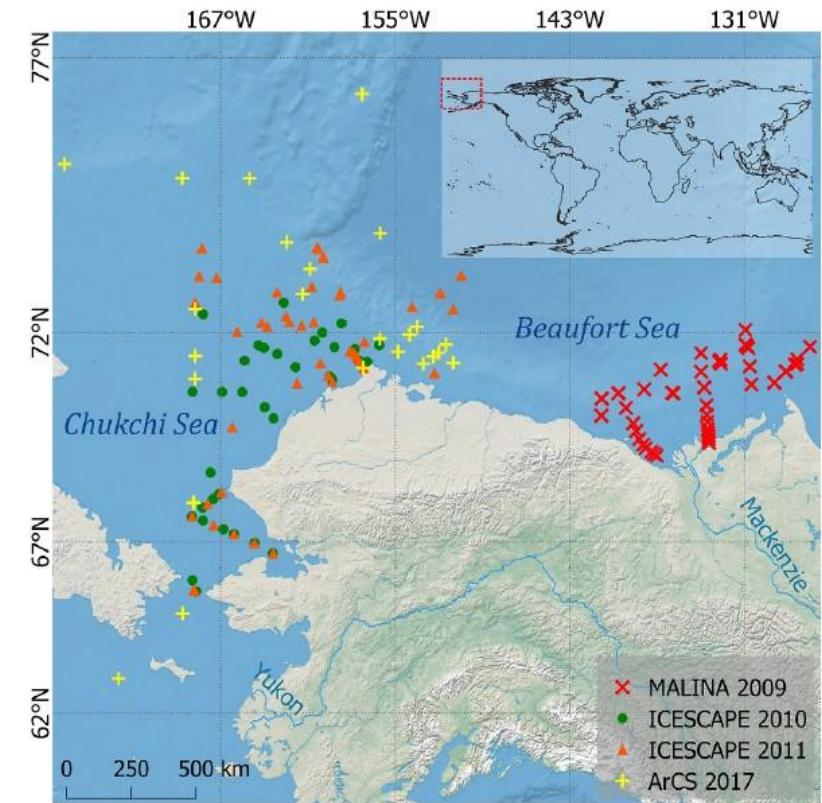
### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif

Etude de Stramski et al. (2023)

*Adaptive optical algorithms with differentiation of water bodies based on varying composition of suspended particulate matter: A case study for estimating the particulate organic carbon concentration in the western Arctic seas*

- 4 campagnes océanographiques
- 335 mesures concomitantes de POC, SPM,  $a_p(\lambda)$ ,  $b_{bp}(\lambda)$ ,  $R_{rs}(\lambda)$
- Large gamme d'environnements optiques/biogéochimiques
- Chlorophylle-a = [0.1 – 30 mg m<sup>-3</sup>]
- POC = [7 – 1750 mg m<sup>-3</sup>]
- SPM = [ 20 – 20 000 mg m<sup>-3</sup>]
- Assemblage particulaire = [totalement minéral – totalement organique]



**Figure 20.** Localisation des mesures concomitantes *in situ* de propriétés optiques inhérente,  $IOP(\lambda)$ , de reflectance,  $R_{rs}(\lambda)$  et de carbone organique particulaire (POC) dans la région de l'océan Arctique (Stramski et al. 2023).

### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif

1<sup>er</sup> étape : développer un algorithme permettant d'estimer le rapport POC/SPM à partir des IOPs

Régression linéaire multiple entre le rapport POC/SPM,  $a_p$ , et  $b_{bp}$

$$\log\left(\frac{POC}{SPM}\right) = a_0 + a_1 \times \log\left[\frac{a_p(570)}{a_p(440)}\right] + a_2 \times \log[b_{bp}(550)] + a_3 \times \log\left[\frac{a_p(570)}{a_p(440)}\right] \times \log[b_{bp}(550)]$$

2<sup>nd</sup> étape : Utiliser le rapport POC/SPM pour comme proxy de la composition de la matière particulaire pour discriminer d'un point de vu optique les masses d'eau dominées par les particules organiques, minérales, ou mixte

POC/SPM < 0.12

→ Assemblage à dominance minérale

0.12 < POC/SPM < 0.25

→ Assemblage mixte

POC/SPM > 0.28

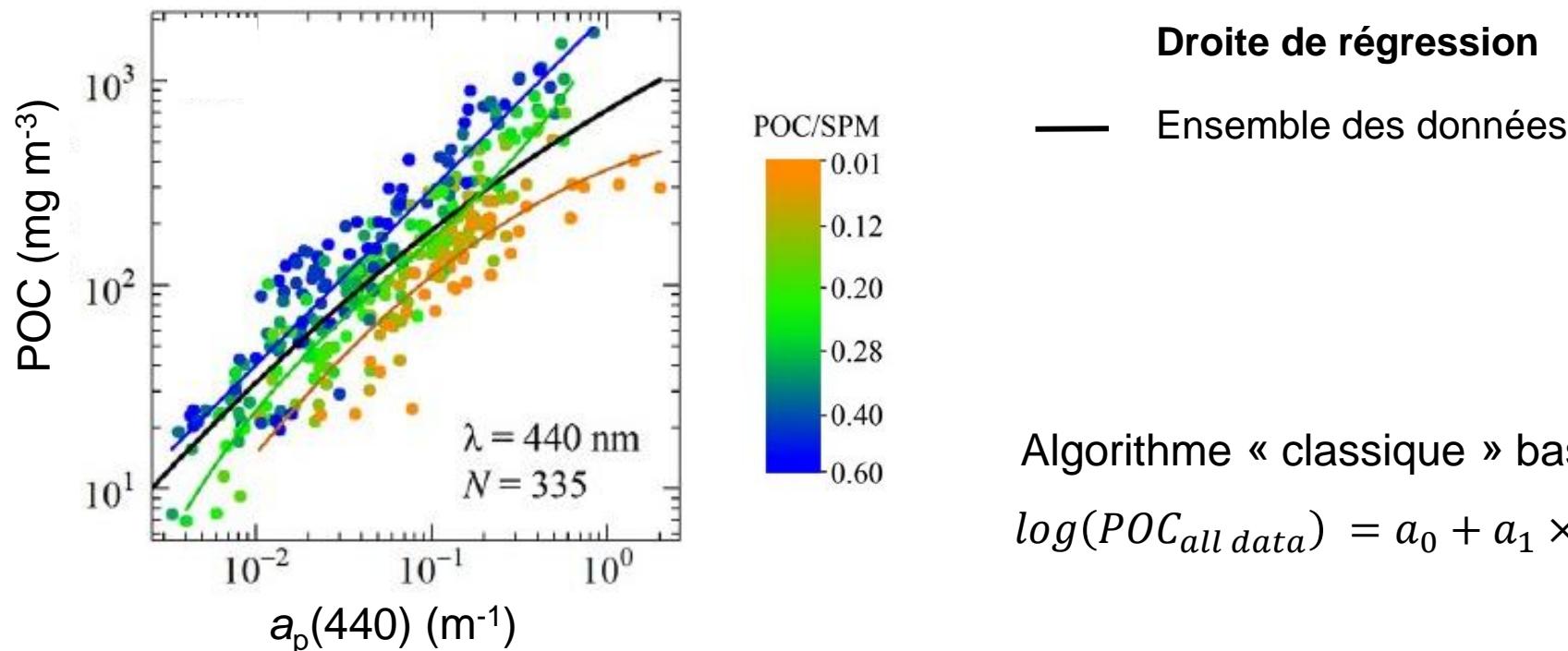
→ Assemblage à dominance organique

Séparation des mesures de POC et  $a_p(440)$  en fonction de ces seuils

### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif

3<sup>em</sup> étape : tests des différentes relations POC vs.  $a_p(440)$



Algorithme « classique » basé sur la relation POC vs.  $a_p(440)$

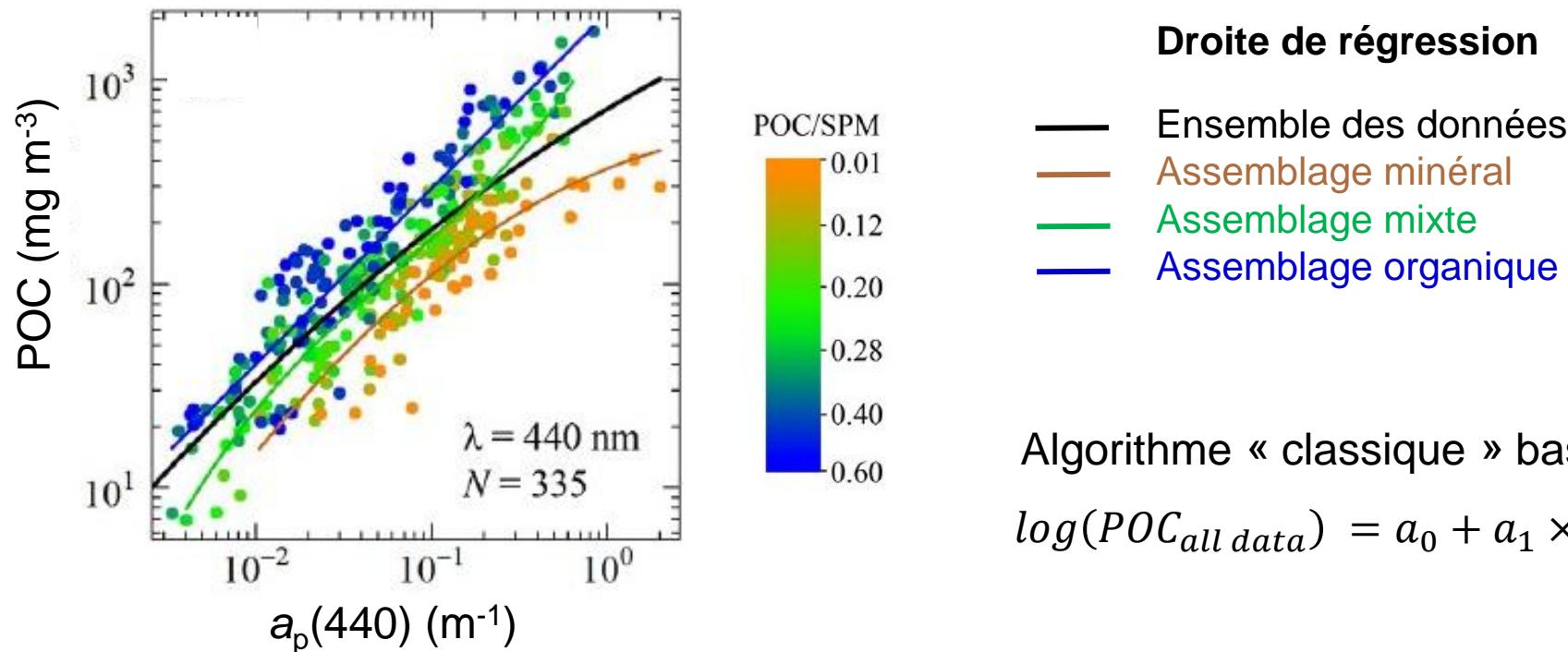
$$\log(POC_{all\ data}) = a_0 + a_1 \times \log[a_p(440) + a_2 \times \log[a_p(440)]^2]$$

**Figure 21.** Relation entre le POC et  $a_p(440)$  en fonction du rapport POC/SPM. Figure tirée de Stramski et al., 2023.

### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif

3<sup>em</sup> étape : tests des différentes relations POC vs.  $a_p(440)$



Algorithme « classique » basé sur la relation POC vs.  $a_p(440)$

$$\log(POC_{all\ data}) = a_0 + a_1 \times \log[a_p(440)] + a_2 \times \log[a_p(440)]^2$$

Algorithme adaptatif en fonction de la composition de l'assemblage particulaire

$$\log(POC_{mineral}) = b_0 + b_1 \times \log[a_p(440)] + b_2 \times \log[a_p(440)]^2$$

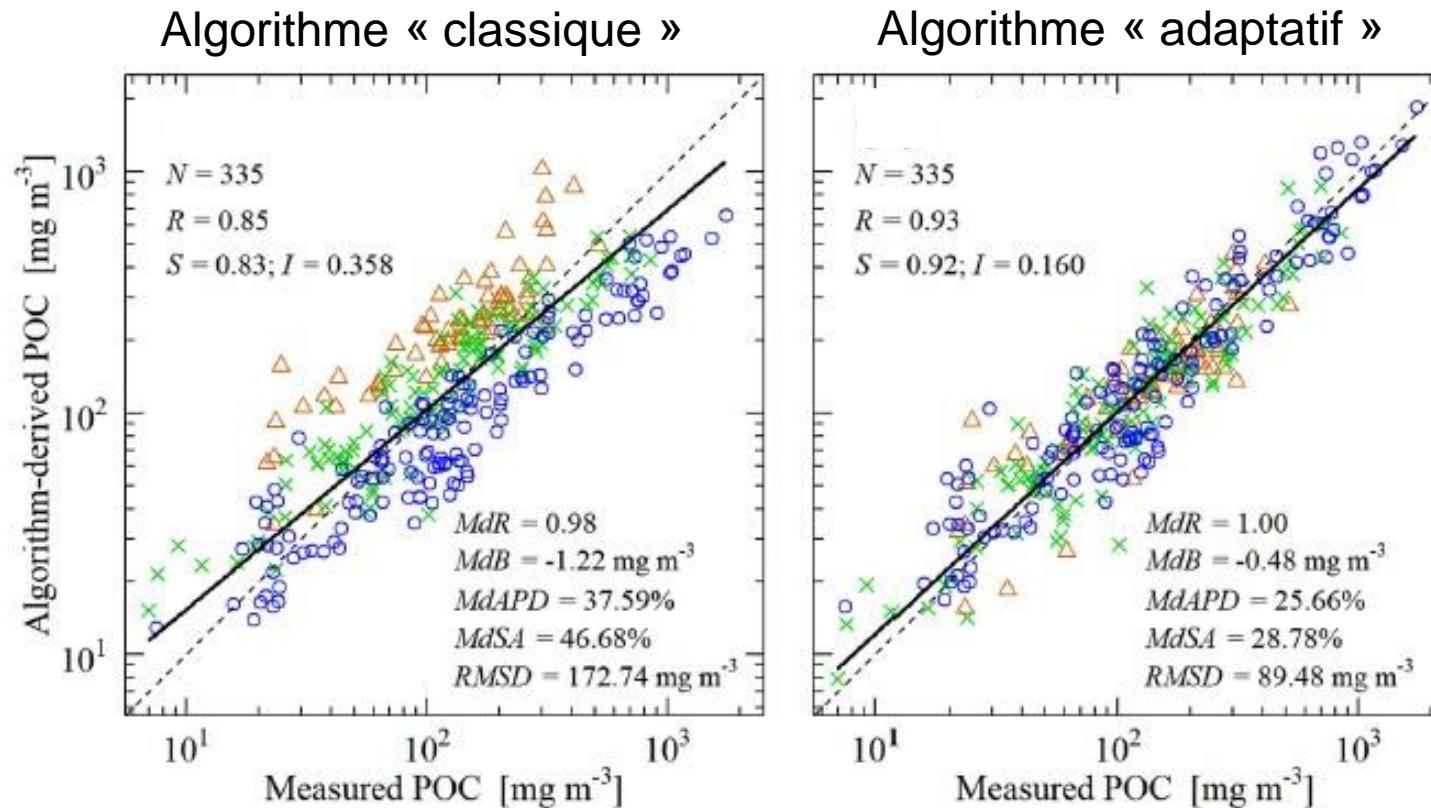
$$\log(POC_{mixte}) = c_0 + c_1 \times \log[a_p(440)] + c_2 \times \log[a_p(440)]^2 + c_3 \times \log[a_p(440)]^3$$

$$\log(POC_{organique}) = d_0 + d_1 \times \log[a_p(440)]$$

### 3) Cas d'étude : le carbone organique particulaire

#### 3.6) Algorithme adaptatif

4<sup>em</sup> étape : comparaison de la performance de l'algorithme « classique » vs. « adaptatif »



La prise en compte de la variabilité de l'assemblage particulaire permet une amélioration notable de l'estimation du POC à partir de mesures optiques

- △ Assemblage minéral
- ✗ Assemblage mixte
- Assemblage organique

**Figure 22.** Relation entre le POC dérivé de l'algorithme classique (gauche) ou adaptatif (droite) et le POC mesuré. Figure tirée de Stramski et al., 2023.

## 4) Résumé

- ❖ Il existe plusieurs types d'algorithmes permettant d'estimer des variables biogéochimiques à partir de mesures optiques : empiriques vs. semi-analytiques.
- ❖ Les différentes étapes permettant le développement et la validation d'un algorithme
  - 1) Assembler une base de données *in situ* du paramètre biogéochimique que l'on cherche à estimer
  - 2) Faire un contrôle qualité rigoureux du jeu de données *in situ*
  - 3) Base de données de *matchups* (NASA Ocean Color protocol, Bailey and Wang, 2001)
  - 4) Choix du type de modèle de régression
  - 5) Validation de l'algorithme, à partir d'un jeu de données indépendant, et de paramètres statistiques.  
En supplément d'un jeu de données indépendant, il est recommandé d'utiliser des données provenant des mesures réalisées au niveau des séries temporelles (station ALOHA, BATS, réseau SOMLIT), ce qui permet de tester la capacité d'un algorithme à reproduire la variabilité temporelle d'une variable biogéochimique.
- ❖ La variabilité des mesures optiques dépend la concentration totale de matières particulaires en suspension, de la composition de l'assemblage particulaire (minéral, organique, type de phytoplancton, contribution du vivant vs. détritique...) et de la taille des particules.
- ❖ Par conséquent, la formulation et la performance des algorithmes sont dépendants du jeu de données (taille et précision du jeu de données *in situ*, couverture spatiale et saisonnière des mesures, changements environnementaux...) avec lesquels ils sont développés.