

PRODUITS BIOGÉOCHIMIQUES 3D DÉRIVÉS D'OBSERVATIONS SATELLITAIRES & IN SITU À L'AIDE DU MACHINE LEARNING (SOCA)

Raphaëlle Sauzède, Julia Uitz,
Louis Terrats, Renosh PR, Hervé Claustre





Les **communautés phytoplanctoniques**, de par leur **biomasse** et leur **grande diversité**, jouent un rôle fondamental dans des processus critiques : (1) la **capacité de l'océan à piéger du carbone** d'origine atmosphérique; (2) la **structuration des réseaux trophiques** (chaînes alimentaires)

Développer une vision 3D de la **biomasse** et de la **composition** des communautés phytoplanctoniques dans l'océan représente donc un défi majeur dont les enjeux sont à la fois climatique et socio-économique

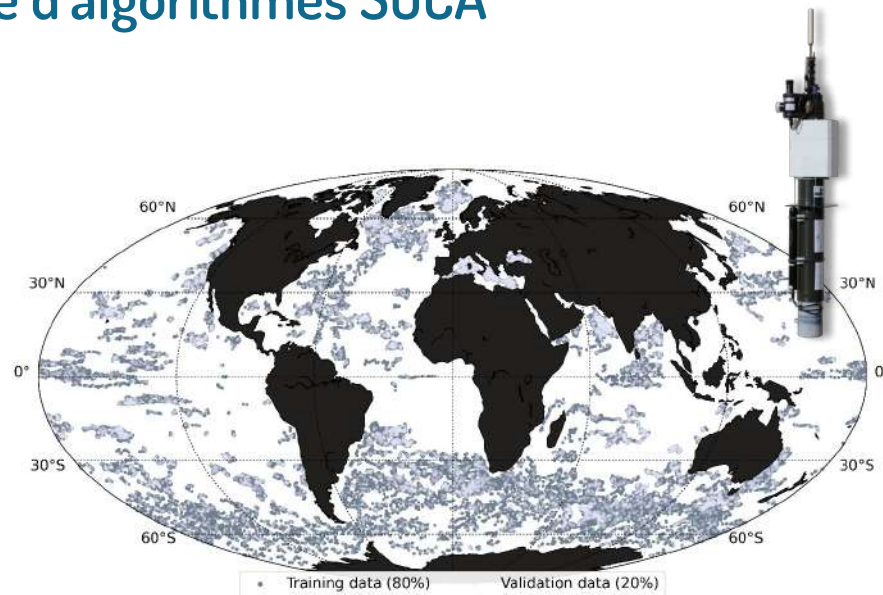
A l'heure actuelle, la **plupart des algorithmes** basés sur les **observations satellitaires OC** donnent une information intéressante mais **limitée à la couche de surface** des océans

Ici nous proposons une famille d'algorithmes (SOCA) permettant de dériver des **produits biogéochimiques 3D**

Contexte de développement de la famille d'algorithmes SOCA

Le réseau d'observation global **BGC-Argo** s'est fortement densifié au cours des 10 dernières années et fournit aujourd'hui des **bases de données bio-optiques riches**, permettant de développer des produits basés sur des méthodes de machine learning (gourmandes en données d'apprentissage/validation)

L'approche proposée tire profit de la **couverture synoptique des observations satellitaires** (OC & SLA) **couplée aux mesures bio-optiques BGC-Argo résolues sur la verticale**, pour développer les algorithmes SOCA, basés sur des réseaux de neurones artificiels (ANN), qui infèrent la distribution verticale de différents produits : Chla, POC (b_{bp}), variables radiométriques (PAR)

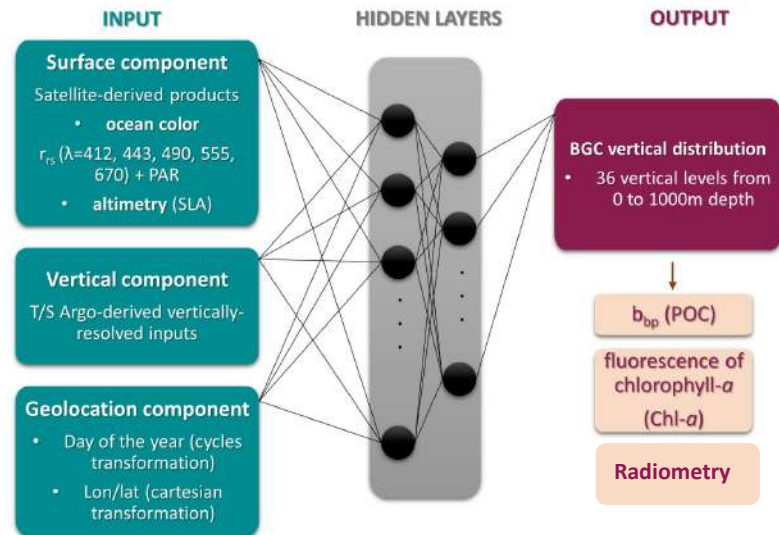


→ Base de données de ~55,000 stations matchups satellite/BGC-Argo utilisées pour l'apprentissage et la validation de la famille d'algorithmes SOCA

Le principe de la famille d'algorithmes SOCA

Le réseau d'observation global **BGC-Argo** s'est fortement densifié au cours des 10 dernières années et fournit aujourd'hui des **bases de données bio-optiques riches**, permettant de développer des produits basés sur des méthodes de machine learning (gourmandes en données d'apprentissage/validation)

L'approche proposée tire profit de la **couverture synoptique des observations satellitaires** (OC & SLA) **couplée aux mesures bio-optiques BGC-Argo résolues sur la verticale**, pour développer les algorithmes SOCA, basés sur des réseaux de neurones artificiels (ANN), qui infèrent la distribution verticale de différents produits : Chla, POC (b_{bp}), variables radiométriques (PAR)



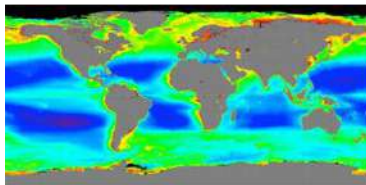
→ Principe de l'architecture de SOCA

Chla & POC : Sauzède et al. in prep.; updated from Sauzède et al., 2016

Radiométrie : Renosh et al. 2023

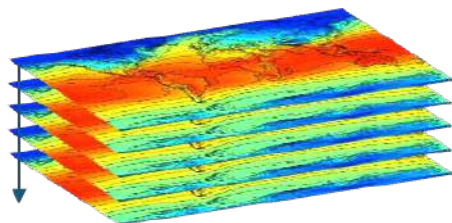
Production de produits globaux 3D (4D) de Chla, POC...

**COPERNICUS SATELLITE FIELDS
(GLOBCOLOUR AND DUACS)**

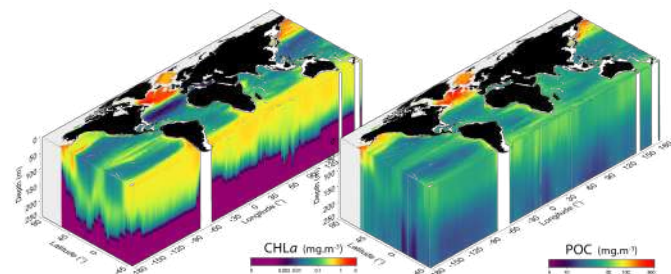


+

**COPERNICUS ARMOR 3D FIELDS
(T/S, MLD)**



GLOBAL 3D OUTPUT PRODUCTS



Produits opérationnels distribués par Copernicus Marine Service (CMEMS)

- Produit [MULTIOBS_GLO_BIO_BGC_3D_REP_015_010](#) issu du MULTIOBS TAC
- 0.25°x0.25° (résolution horizontale)
- 36 niveaux verticaux de la surface à 1000 m de profondeur (résolution verticale)
- Champs hebdomadaires de 1998 à 2021 (résolution temporelle)
- Champs climatologiques mensuels (résolution temporelle)

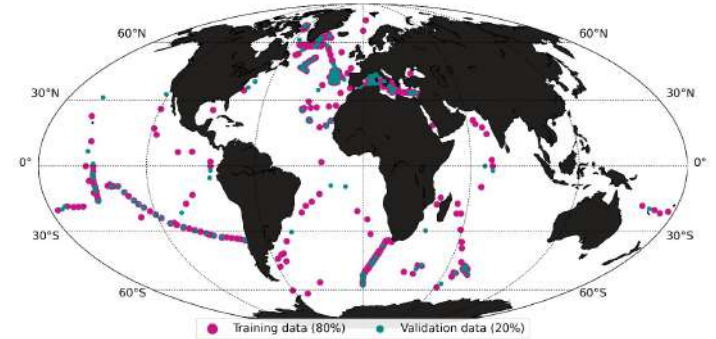
SOCA-PFT: une étape supplémentaire → l'algorithme FLAVOR-TS

Nous proposons ici de **développer un algorithme SOCA-PFT** suivant le même principe que SOCA-Chla ou SOCA- b_{bp}

Cela nécessite une étape additionnelle permettant **d'enrichir le jeu de données d'apprentissage matchup satellite/BGC-Argo de la variable PFTs**, actuellement non mesurée par les flotteurs BGC-Argo

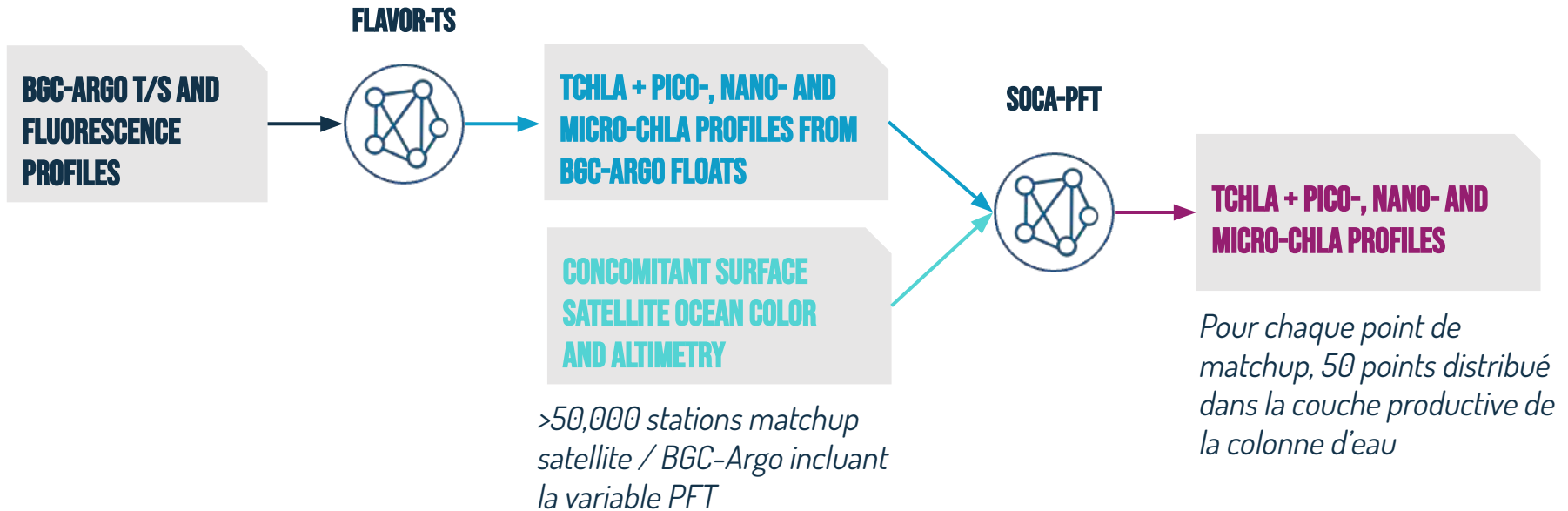
Nous avons construit une **base de données** de profils verticaux de **fluorescence + T/S** - tels que mesurés par les flotteurs BGC-Argo - et des **pigments** déterminés par HPLC (mesures de référence TChla et composition phytoplanctonique)

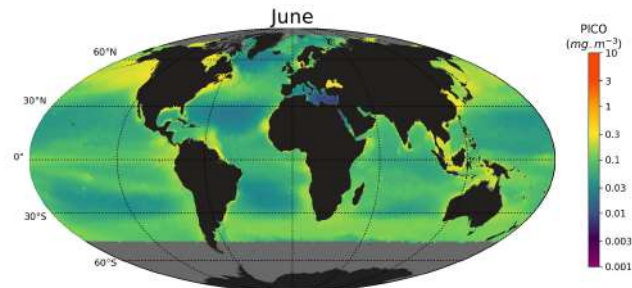
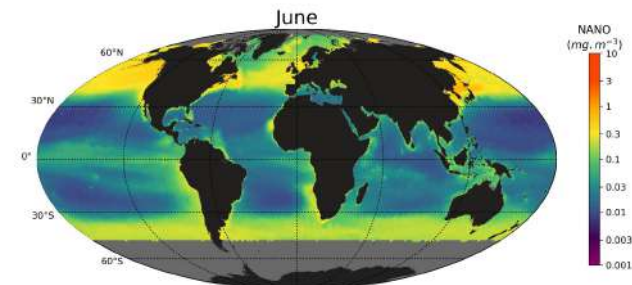
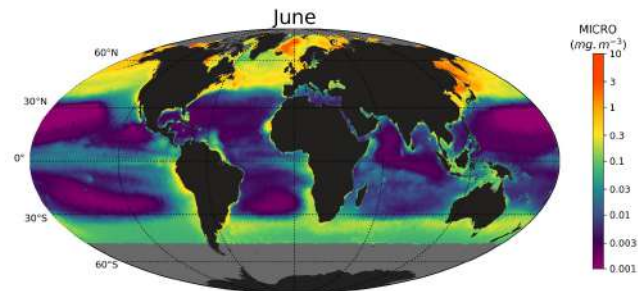
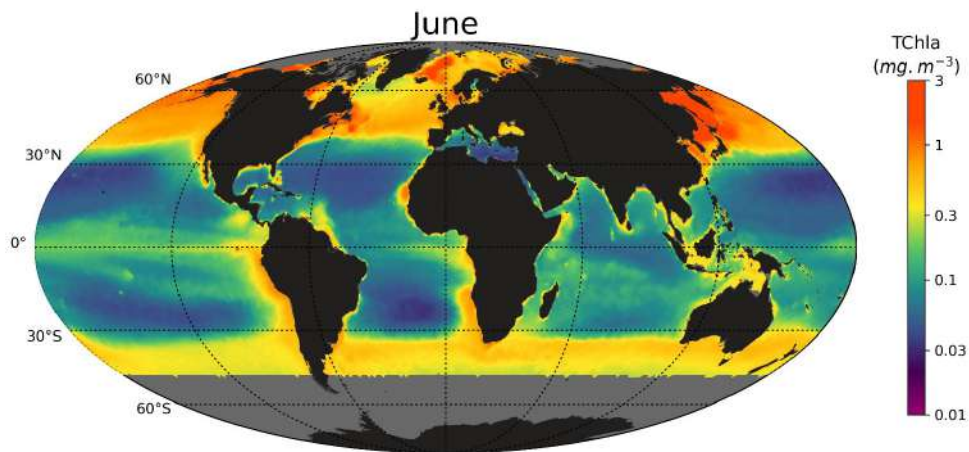
Cette base de données est utilisée pour **entraîner/valider FLAVOR-TS** (version upgraded de FLAVOR, Sauzède et al. 2015).



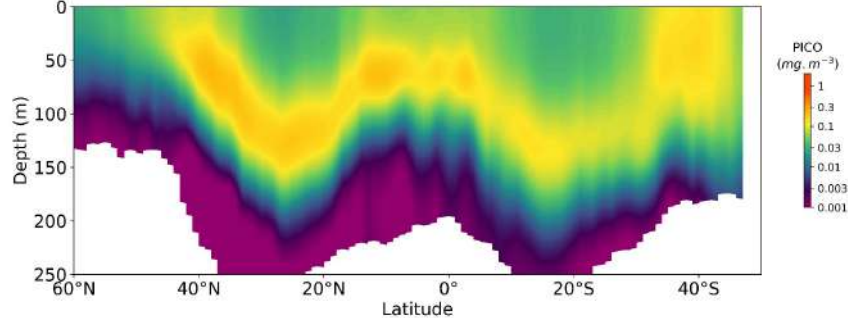
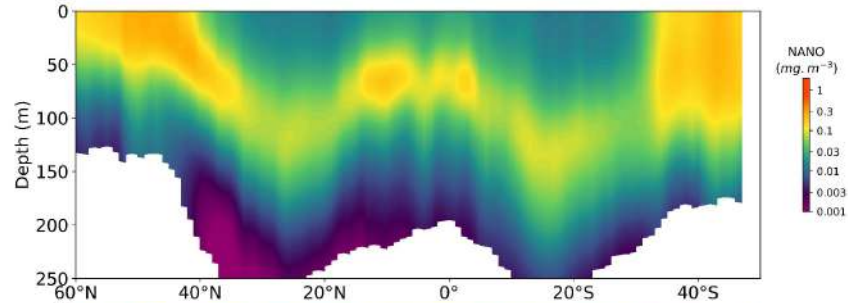
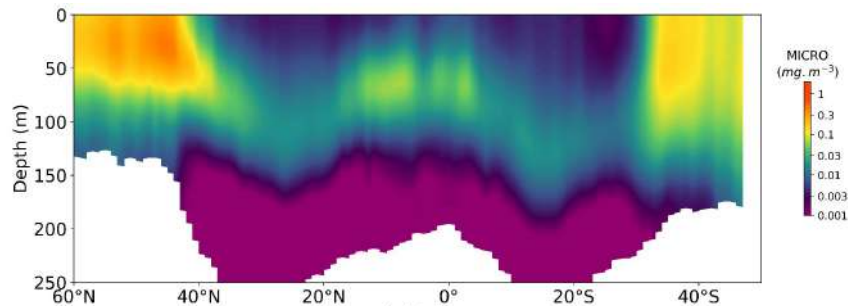
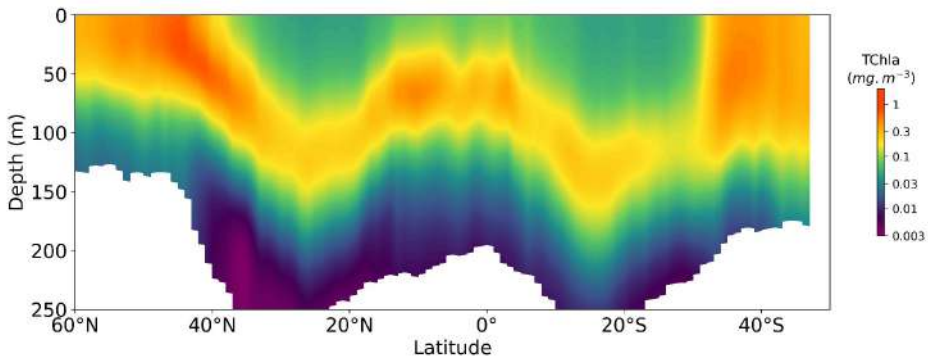
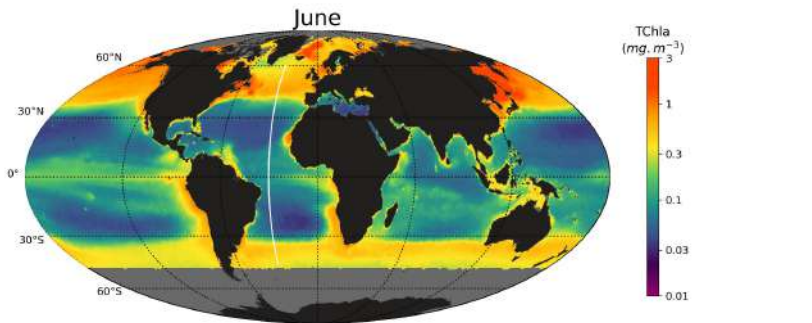
→ Localisation des ~1600 profils verticaux de fluorescence, T/S et pigments HPLC utilisés pour l'apprentissage/validation de la méthode FLAVOR-TS

Implémentation de la méthode à 2 étapes : FLAVOR-TS + SOCA-PFT





Exemple d'un transect Atlantique en Juin (climato)



Contexte & objectifs

Projet à l'**interface atmosphère/océan**, porté par Karine Sellegri au LaMP

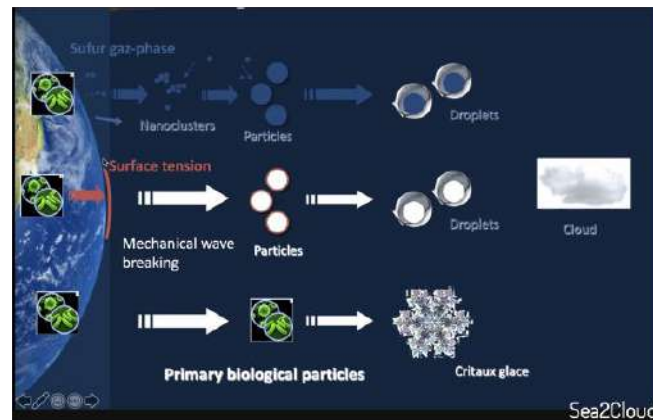
Etudier le **rôle des microorganismes marins sur la formation et les propriétés des nuages**, via différents mécanismes encore méconnus et négligés dans les modèles, avec un focus sur l'Océan Austral peu documenté

Outils principaux : Campagne de terrain, **produits satellites large échelle**, modélisation

Rôle du LOV :

(1) Apporter une expertise sur la biogéochimie et les communautés phytoplanctoniques

(2) Fournir des produits satellitaires relatifs au phytoplancton (Chla, POC, PFTs)



Réalisations au LOV : 12 mois de CDD IR CNES + 3 mois reliquats BlueCloud

- ✓ **Développement du produit de PFTs global quasi-finalisé, incluant**
 - **Intégration de nouvelles données** : HPLC training de FLAVOR et matchup satellite/BGC-Argo training SOCA-PFT
 - Prise en compte de nouveaux pigments pour une **résolution taxonomique plus fine**
 - Amélioration de la **grille de profondeurs** sur laquelle sont résolus les produits (distribution et profondeur max)
 - **Score statistique** multicritères pour **optimisation** du choix des **architectures** et de la taille des **ensembles**
 - Multiples tests non concluant visant à améliorer les résultats (ex. transformation des données en ACP, etc.)

- ✗ **Développement d'un produit de PFTs Austral**
 - Non réalisé

- ✓ **Collaboration avec le LaMP**
 - Distribution des produits SOCA-Chla + SOCA-POC (CMEMS) et ancien produit PFT (Uitz et al 2006)
 - Publications : Sellegri et al. 2021 (Sci Report), Bazantay et al. soumis (GRL), Rocco et al. soumis (JGR-Atm)

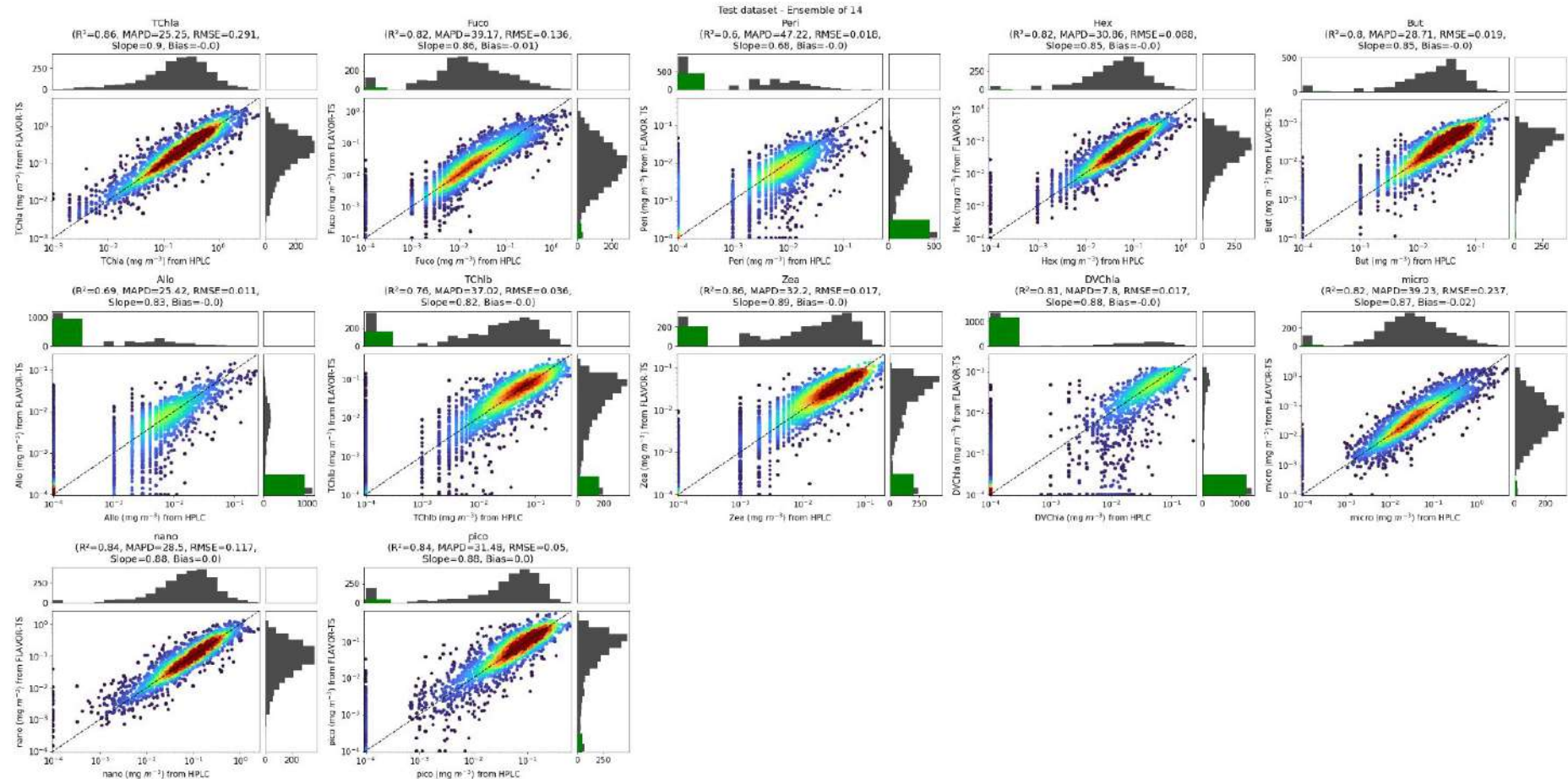
Réalisations au LOV : 12 mois de CDD IR CNES + 3 mois reliquats BlueCloud

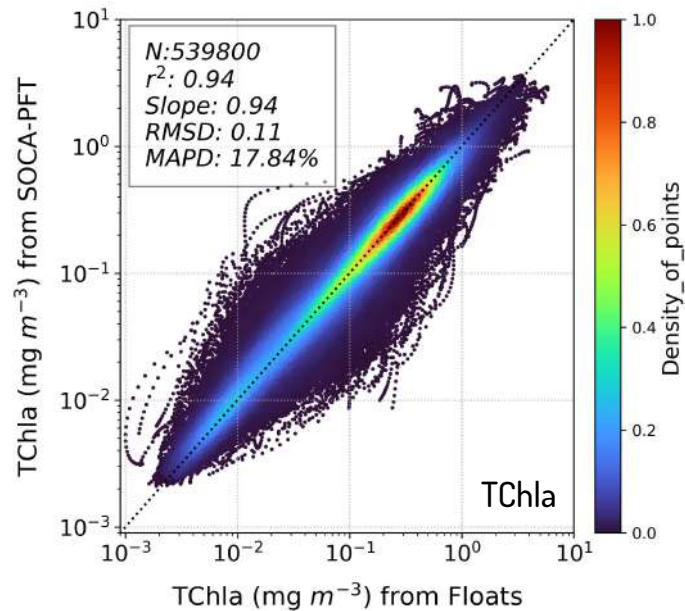
Perspectives court (3 semaines) et moyen terme

- Finalisation de la validation basée sur données de référence HPLC (chaîne complète FLAVOR-TS → SOCA-PFT et comparaison aux anciens modèles)
- Evaluation du nouveau modèle pour la région Australe vs autres régions
- Développement de projets sur l'Austral qui reste un point central indépendamment de Sea2Cloud (cf. postdoc de Nicolas Mayot)

Merci de votre attention!







- Erreur moyenne faible : 18% pour la TChla, 20-27% pour les groupes phytoplanctoniques

- La méthode représente bien le gradient de biomasse de la TChla et de chacun des PFTs

