



## Pôle de Calcul et de Données pour la Mer Rapport d'activité 2018



UNION EUROPÉENNE  
UNANIEZH EUROPA



L'Europe s'engage  
en Bretagne / Avec le Fonds européen  
de développement régional





**Titre du rapport : Rapport d'activités 2018 du Pôle de Calcul et de Données pour la Mer (PCDM)**

**Diffusion :**

libre (internet)

restreinte (intranet) – date de levée d'embargo : AAA/MM/JJ

interdite (confidentielle) – date de levée de confidentialité : AAA/MM/JJ

**Date de publication :** 01/07/2019

**Version :** 1.0.0

**Référence de l'illustration de couverture**

@Ifremer

**Langue(s) :** Français

**Projets dans lesquels ce rapport s'inscrit** (programme européen, campagne, etc.) : CPER Datarmor

**Auteur(s) / adresse mail**

**Affiliation / Direction / Service, laboratoire**

Pierre Cotty - Pierre.Cotty@ifremer.fr

PDG-IRSI

Anne Thomas-Doaré - Anne.Thomas.Doare@ifremer.fr

PDG-IRSI-RIC

Denis Croizé-Fillon - Denis.Croize.Fillon@ifremer.fr

PDG-IRSI-RIC

Avec la contribution des utilisateurs du PCDM

**Validé par :** [Pierre.Cotty@ifremer.fr](mailto:Pierre.Cotty@ifremer.fr)

# Table des matières

1	Introduction.....	7
2	Le contexte « données marines ».....	7
2.1	L'océan, un milieu partagé .....	7
2.2	Le centre de données SISMER .....	8
2.3	Le pôle océan ODATIS au niveau national .....	8
2.4	Le paysage européen des données marines.....	9
2.5	Les enjeux.....	9
3	Gouvernance et comité de pilotage.....	11
4	Organisation et statistiques .....	12
4.3	Statistiques de traitements/calculs .....	15
4.3.1	Temps CPU consommé.....	16
4.3.2	Consommation processeur.....	16
4.3.3	Nombre de jobs .....	18
4.3.5	Statistiques par unité .....	20
4.3.6	Nombre d'utilisateurs .....	21
4.4	Statistiques logistiques .....	21
4.5	Disponibilité.....	22
5	Services et Ingénierie du PCDM.....	23
5.1	Upgrade 2018 dans le cadre du CPER .....	23
5.2	La bioinformatique.....	23
5.2.1	La mise à disposition des logiciels .....	23
5.2.2	Les applications de bio-informatique .....	24
5.2.3	La bancarisation des données de bioinformatique.....	25
5.3	Organisation des Web services .....	25
5.3.1	Virtualisation d'applications Web sur la grappe ClusterWeb .....	25
5.3.2	Plate-forme web JupyterHub .....	26
5.3.3	Plate-forme Elastic Search.....	26
5.4	Communication/Informations .....	27
5.4.1	Journée des utilisateurs 2018 .....	27
5.4.2	Documentation / Site Web PCDM .....	28
5.4.3	Enquête utilisateurs .....	29

6	Les exemples d'utilisation du PCDM.....	36
6.1	Recherche académique .....	36
6.1.1	L'exploitation de l'imagerie hyperspectrale pour la caractérisation de la végétation marine : une première sur Datarmor .....	36
	<i>Touria BAJJOUK – Ifremer DYNECO/LEBCO</i> .....	36
6.1.2	Etude de l'influence des forçages naturels et anthropiques sur la dynamique hydro-sédimentaire en Estuaire de Seine.....	39
6.1.3	Influence des tourbillons sur la variabilité intrinsèque basse fréquence et grande échelle de l'Atlantique Nord.....	40
6.1.4	Modélisation haute résolution de la convection profonde en mer Méditerranée Nord-Occidentale.....	41
6.1.5	ETude des Ondes Internes au large des LandEs (ETOILE) .....	43
6.1.6	Régionalisation des modèles de climat dans le Pacifique Sud .....	46
6.1.7	Mise en œuvre et performances du modèle CROCO pour les configurations cotières.....	49
6.1.8	Projet DIETE .....	50
	<i>Martin Plus – DYNECO Pélagos</i> .....	50
6.1.9	Projet Alex-Breizh .....	51
6.1.10	Rôle de l'adaptation locale et de la connectivité dans le structure et le fonctionnement des populations de coquilles Saint Jacques ( <i>Pecten maximus</i> ) en Rade de Brest, Mer d'Iroise et Manche.....	51
6.1.11	Projet FOREVER : Modélisation de la dispersion larvaire en rade de Brest .....	53
6.1.12	OSMOSE-MED : un modèle écosystémique intégré pour comprendre et gérer les ressources d'une Méditerranée sous pressions .....	56
6.1.13	Le modèle APECOSM.....	56
6.1.14	Divers travaux de l'unité MARBEC .....	57
6.1.15	Projet PRESENCE : PRESsions sur les Ecosystèmes récifo-lagonaires de Nouvelle-Calédonie .....	59
6.1.16	Étude de sensibilité de modèles hydrologiques pour la prévision des crues dans un estuaire.....	60
6.1.17	Étude de données de vent en Europe pour caractériser le potentiel de l'éolien en mer.....	62
6.1.18	Optimisation du frottement sur le fond en environnement macrotidal : application aux courants barotropes du Raz Blanchard .....	65
6.2	Entreprises.....	69
6.2.1	BV Solutions M&O : Cas test de slamming de maxi-trimaran de course au large	69
6.2.2	BV Solutions M&O – RealTide Project .....	71
6.2.3	BV Solutions M&O : Optiroutes .....	74

6.2.4	EXWECS .....	77
6.3	Applications opérationnelles .....	78
6.3.1	Centre Aval de Traitement des Données SMOS – CATDS .....	78
6.3.2	MARC .....	79
6.3.3	Centre opérationnel de données satellite CERSAT .....	87
7	Configuration technique de DATARMOR .....	90
7.1	Architecture générale.....	90
7.2	Les clusters d'exécution.....	91
7.3	Les clusters de stockage.....	93
7.4	Les interfaces et les réseaux.....	95
7.5	La logistique .....	96
7.6	Les liaisons réseau inter-organisme .....	98
7.7	Les principaux logiciels .....	98
8	Annexes.....	100
8.1	Liste des membres du Comité de Pilotage Datarmor .....	100
8.2	Glossaire .....	101

# 1 Introduction

Depuis 2001, le Pôle de Calcul et de Données pour la Mer (PCDM) opère une infrastructure informatique mutualisée entre l'IFREMER, le SHOM, l'UBO (dont l'IUEM), l'ENSTA Bretagne, l'Ecole Navale, le CEREMA et des PME. Accessible à toute la communauté des Sciences Marines, l'infrastructure numérique mutualisée est hébergée sur le site IFREMER de Brest/Technopôle.

Afin de renouveler et adapter les infrastructures techniques aux nouveaux besoins des sciences marines, le projet Datarmor a donc été construit dans le cadre du CPER Bretagne 2015-2020, et a pu être financé. Au regard de la configuration technique de Datarmor et du panel de services rendus aux utilisateurs, désormais très orientés « données », le COPIL a également opté pour un changement de nom du consortium, afin de mieux coller aux réalités techniques : PCIM (Pôle de Calcul Intensif pour la Mer) est devenu PCDM (Pôle de Calcul et de Données pour la Mer).

Le présent rapport d'activités porte sur l'année 2018, première année pleine d'exploitation de Datarmor.

## 2 Le contexte « données marines »

### 2.1 L'océan, un milieu partagé

Les sciences marines se positionnent particulièrement bien dans la tendance lourde que constitue pour les sciences en général l'approche « systémique » ou « écosystémique » ; en effet l'océan représente un milieu particulièrement intégré, en mouvement, où toutes les disciplines doivent interagir.

De ce fait, les systèmes d'observation, les bases de données, les outils de modélisation peuvent être, dans une large mesure, fédérés et mutualisés autour des sciences marines, et même un peu plus largement avec les milieux connexes que sont l'atmosphère, la terre solide et les surfaces continentales qui constituent avec l'océan les 4 pôles thématiques de l'Infrastructure de Recherche (IR) « système Terre » DATA TERRA, labellisée par le Ministère de la Recherche (MESRI)

C'est particulièrement le cas pour les outils informatiques afférents, qui doivent cumuler des capacités importantes de calcul/traitement avec des accès rapides à de grosses quantités de données.

En tant qu'institut de recherche spécialisé en sciences marines, Ifremer a vocation à poursuivre une stratégie de fédération de l'ensemble des données marines et des outils de traitement/calcul nécessaires à leur interprétation ou modélisation ; cette stratégie s'inscrit logiquement au niveau national dans une contribution importante à l'IR « Data Terra » et son pôle océan ODATIS, et dans les projets européens qui portent la même démarche de fédération.

## 2.2 Le centre de données SISMER

Dès 1961, la COI (Commission Océanographique Internationale) crée un programme appelé IODE (Intergovernmental Oceanographic Data Exchange) pour promouvoir la mise en commun et l'ouverture des données, bien avant l'avènement des politiques « open data / open sciences »

IODE se développe ensuite et définit une organisation notamment basée sur des NODC (National Oceanographic Data Center).

Le concept existe toujours et IODE attribue une certification pour les NODC respectant un panel de critères précis.

Le SISMER est certifié en tant que centre de données « NODC » par la COI. Il est également partie prenante au niveau national du pôle océan ODATIS, et coordinateur de projets européens sur les données marines. Le SISMER est hébergé à Ifremer Brest et est utilisateur de l'infrastructure Datarmor pour le stockage, le traitement et la diffusion de données marines, notamment les plus volumineuses.

Les données opérées par le SISMER se divisent en grandes filières :

- Les données satellite
- Les données d'océanographie « in situ »
- Les données halieutiques
- Les données d'environnement littoral
- Les données de biodiversité
- Les données issues des campagnes de la Flotte Océanographique

## 2.3 Le pôle océan ODATIS au niveau national

Apparue en 2016, puis confirmée en 2018 dans la feuille de route des infrastructures de recherche (IR) du MESRI, l'IR « Pôles de données et de services du système Terre », désormais appelée Data Terra, a pour mission principale de mettre à disposition des données, des produits, des logiciels, des outils et des services destinés en priorité à la communauté scientifique française dans le cadre de ses recherches sur le système terre.

Cette IR comprend 4 pôles qui ciblent chacun un compartiment du système Terre :

- l'atmosphère (pôle AERIS)
- la terre solide (pôle FORMATER)
- les surfaces continentales (pôle THEIA)
- l'océan (pôle ODATIS)

Les 4 pôles s'appuient sur des « centres de données » hébergés dans les établissements partenaires, qui fournissent des services opérationnels de données scientifiques ; le centre de données SISMER, hébergé à Brest est un fournisseur majeur de données du Pôle Océan.



## 2.4 Le paysage européen des données marines

L'UE a bien mis en avant le caractère précieux et prioritaire que revêtent les données scientifiques. Par exemple, il est désormais clairement demandé un plan de gestion de données sur tous les propositions de projets H2020, même celles qui ne s'inscrivent pas directement dans les appels relatifs aux infrastructures de données.

La stratégie UE s'inscrit depuis environ 2 ans sur une politique de « clouds » (1 par DG) ; les services et données produits issus des projets doivent être mis à disposition dans des infrastructures de distribution ciblées (EOSC, DIAS, ...). C'est le cas notamment des projets coordonnés par Ifremer que sont Seadatacloud et Copernicus TAC in situ. En revanche, les processus de « production » de ces jeux de données ou de ces services de données sont répartis entre tous les acteurs, au sein de centres de données. De ce fait, des écueils/difficultés inhérents à un hébergement externalisé et lointain sont esquivés :

- Pas de gestion de SSI compliquée à déléguer à l'hébergeur (les données produites issues des projets ont vocation à être ouvertes, publiques)
- Peu de travail technique à distance de la part des partenaires sur le site de l'hébergeur (l'assemblage des données se fait en amont, près des producteurs de données)
- Pas de nécessité de rebranchement des arrivées de flux de données en provenance des producteurs/systèmes d'observation qui restent connectés sur les sites des partenaires acteurs en amont

Un autre axe important de la stratégie européenne est de promouvoir les FAIR Data (Findable, Accessible, Interoperable, Reusable), c'est-à-dire des interfaces de recherche et consultation des données qui facilitent l'accès. Tous les appels d'offres européens en rapport avec des données scientifiques font référence à cet objectif ; les clouds européens (EOSC, DIAS) qui centralisent les données produites en fin de chaîne ont pour fonction de fédérer et harmoniser les interfaces de distribution (Distribution Units).

Enfin, côté environnement, les directives existantes (Inspire, DCE, DCSMM,...) traduisent l'obligation de porter les données à la connaissance des citoyens, sous forme de systèmes d'informations unifiés en réseau aux niveaux européen et national, systèmes auxquels Ifremer et ses partenaires contribuent via l'échelon national.

## 2.5 Les enjeux

En matière d'infrastructure informatique, les dynamiques de concentration promues par les politiques de cloud tant au niveau national qu'europpéen, poussent à définir de nouveaux principes d'interopérabilité, capable de préserver l'organisation existante des grands systèmes d'observation et des centres de données thématiques, tout en garantissant un accès homogène à l'exhaustivité des données disponibles. Les projets européens de partage de données comme Seadatacloud, Copernicus in situ TAC, Emodnet travaillent sur la définition technique d'un compromis, qui satisfasse au devoir d'un certain niveau de mutualisation informatique porté par les clouds, tout en préservant à chaque communauté de données son autonomie de décision et de fonctionnement.

Sur le plan technique, les performances des traitements de gros volumes de données représentent un challenge très présent, qui doit trouver des réponses autres que la réplication de ces données dans des formats plus adaptés au dit traitement.

Les formats de fichiers, traditionnels en sciences marines, de données géoréférencées comme NETCDF, ne sont probablement plus adaptés à certains traitements, et une urbanisation alternative des données marines reste à mettre au point et à appliquer.

### 3 Gouvernance et comité de pilotage

La finalité de l'infrastructure Datarmor est de nature scientifique ; la machine héberge et contribue aux projets scientifiques en sciences marines, en fournissant des services de :

- Calcul/traitement,
- Accès en mode « services interactifs » ,
- Mise à disposition performante de données de référence,
- Mise à disposition d'espaces de stockage de travail,
- Support technique et thématique sur les applications, les données et les logiciels.

L'ensemble des principes de gouvernance est fixé par des décisions d'un Comité de Pilotage (COFIL), qui s'appuie sur 2 groupes de travail :

- Le groupe de travail « données » (GT données), composé de représentants des principales entités scientifiques, statue notamment sur la sélection de données de référence, et l'attribution d'espaces de stockage « de travail » ,
- Le groupe de travail « entreprises » (GT entreprises) statue sur la liste des entreprises accueillies et les droits d'usages accordés à leurs utilisateurs.

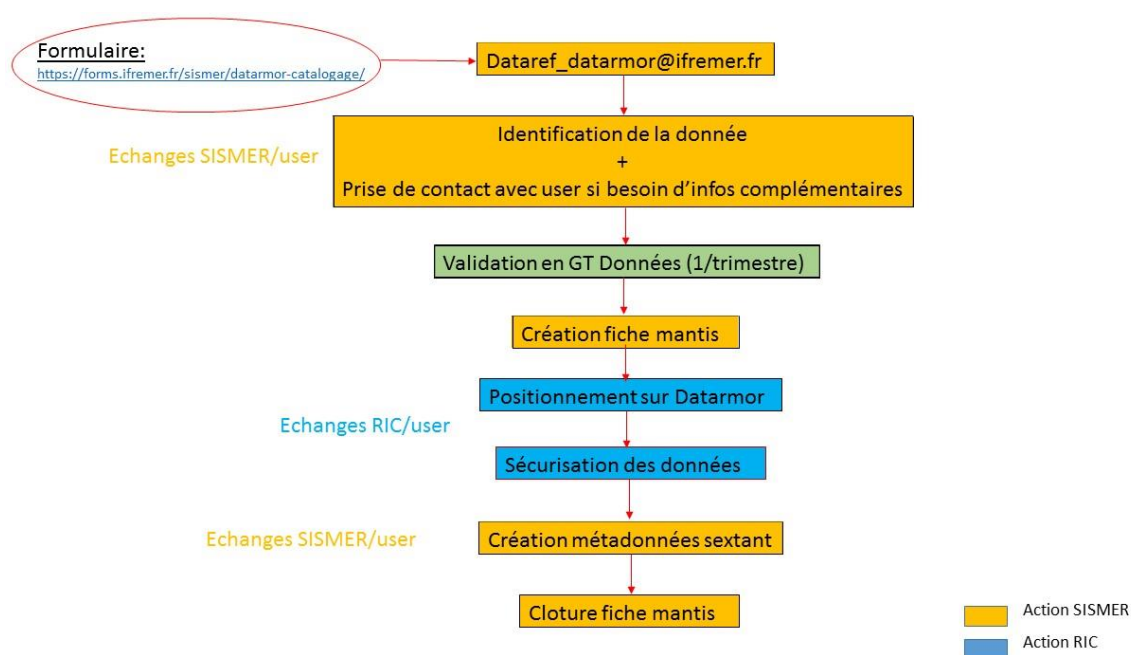
Un rapport d'exploitation Datarmor est édité trimestriellement et envoyé aux membres du comité de pilotage et des groupes de travail.

## 4 Organisation et statistiques

### 4.1 Espace de données de référence

L'espace des données de référence est défini comme un espace de mise à disposition de données d'observation ou non reproductibles (i.e. sans impact significatif en temps et/ou budget) à plusieurs utilisateurs de Datarmor. La volumétrie n'est pas définie comme un paramètre bloquant, mais dimensionnant.

Afin de pouvoir valider une demande d'hébergement de données sur Dataref en GT Données, les utilisateurs doivent remplir le formulaire de catalogage accessible à cette adresse : <https://forms.ifremer.fr/sismer/datarmor-catalogage/>. Le Sismer reçoit alors la demande via la messagerie dédiée [dataref\\_datarmor@ifremer.fr](mailto:dataref_datarmor@ifremer.fr) et déclenche les procédures de récupération et mise à disposition des données en relation avec le service RIC. Le workflow peut être schématisé comme suit :



En 2018, 70 nouvelles demandes d'hébergement sur Dataref ont été formulées. Fin 2018, le nombre total de demandes s'élève ainsi à 213 (dont 53 annulées et 1 non validée). Parmi les communautés les plus consommatrices en terme de nombre de jeux de données hébergés sur dataref, on retrouve :

- L'océanographie opérationnelle côtière qui héberge les jeux de données servant à l'exploitation des modèles MARC
- La bio-informatique
- L'océanographie physique et spatiale

Chaque jeu de données hébergé sur Dataref est décrit via l'infrastructure Sextant accessible directement depuis le site web du [PCDM](#). La description de ces fiches est faite dans le respect

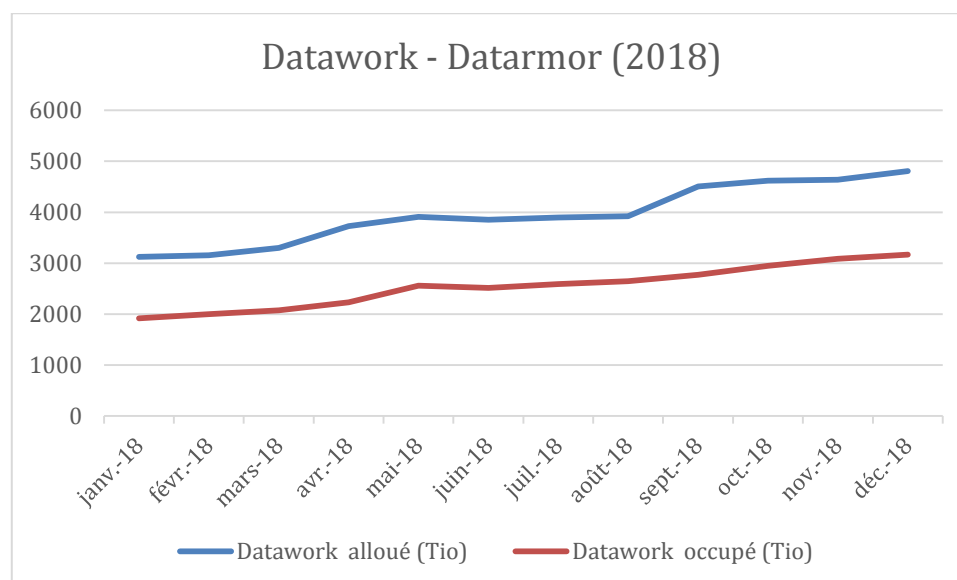
des standards ISO (19115) et de la Directive INSPIRE. En complément des informations relatives aux données, les utilisateurs peuvent également visualiser dans la fiche les protocoles d'accès : chemin sur la machine, ftp, Opendap, formulaire externe, etc. .

## 4.2 Espaces de stockage de travail

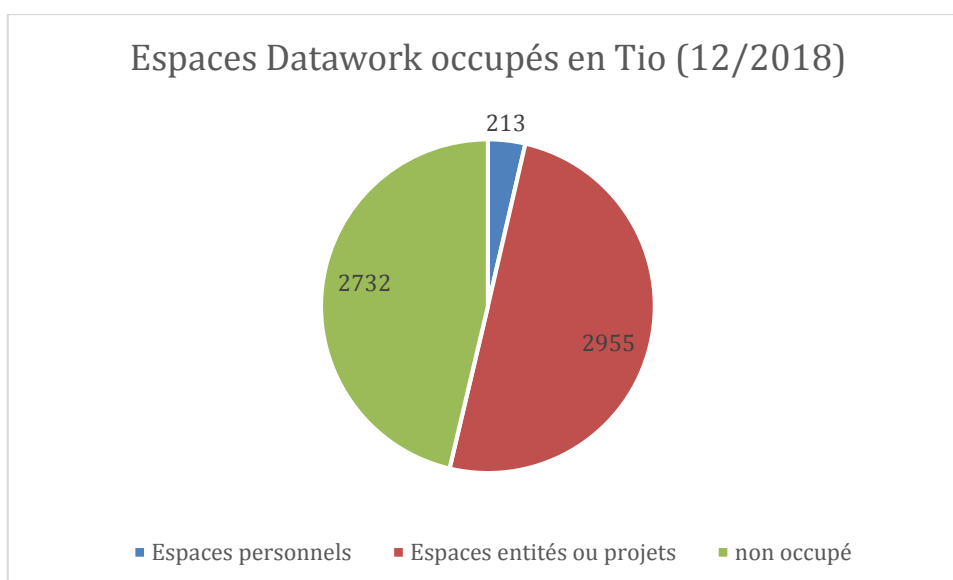
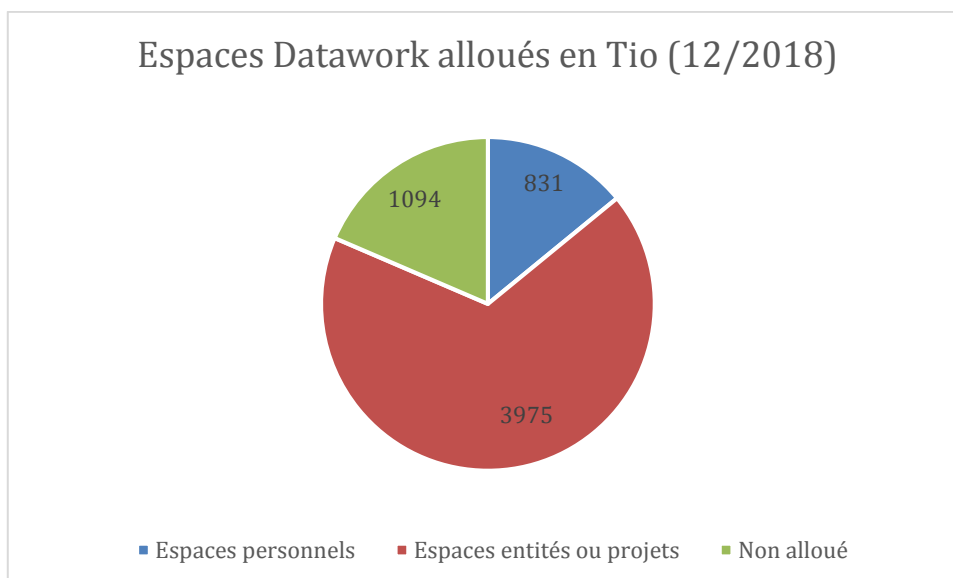
L'espace de travail Datawork est de 5,7 Pio après l'augmentation de 1,2Pio intervenue pendant l'été 2018.

Chaque utilisateur dispose d'un espace de travail personnel de 1Tio. La majeure partie de l'espace est cependant occupée par des espaces communs à des projets ou des équipes. Comme pour les données de référence, l'arbitrage des demandes d'espace est effectué par le groupe de travail « Données ».

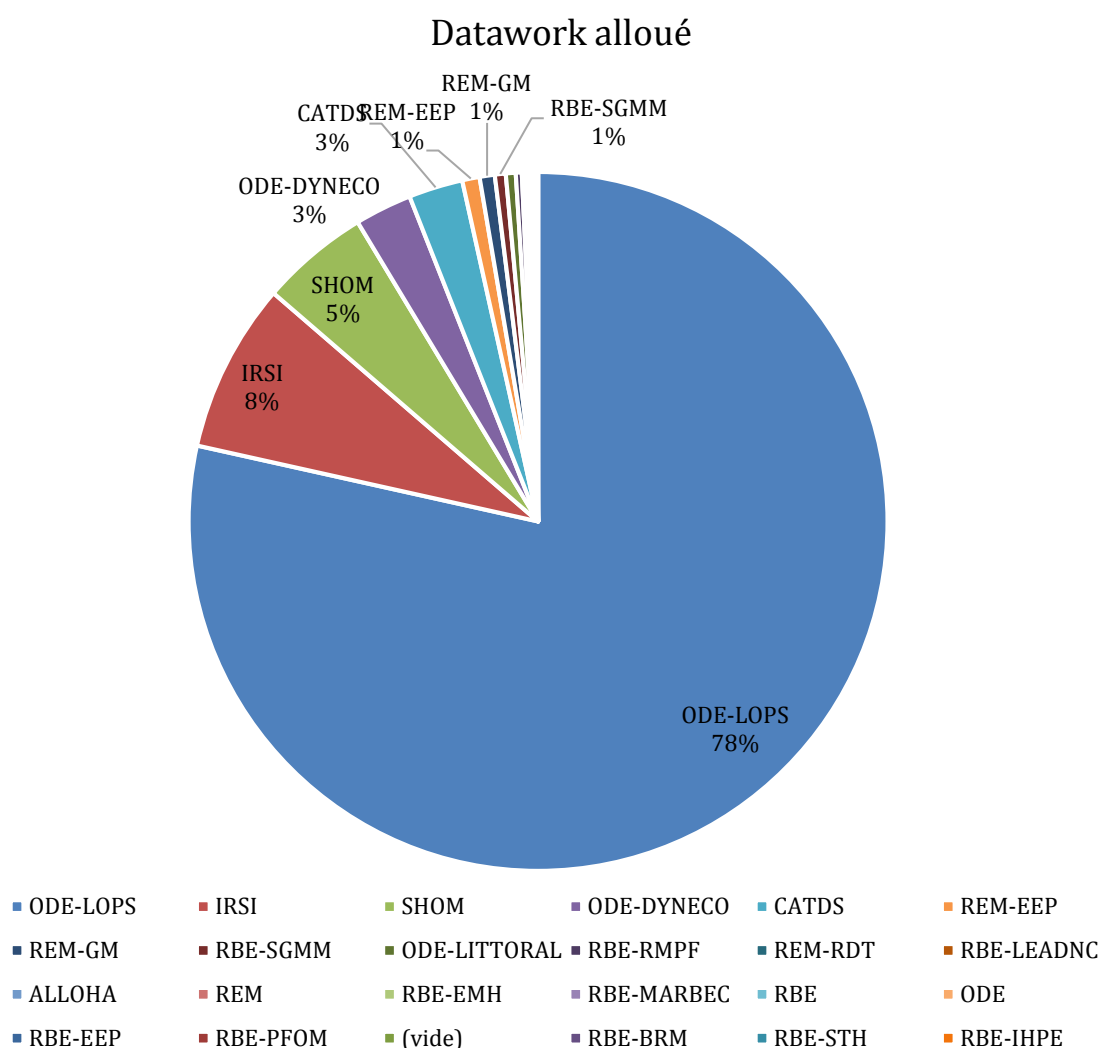
Dans la plupart des cas, quand un utilisateur de Datarmor lance un programme ou une application, il utilise la partie « noeuds de calcul » pour exécuter son code et écrit ses résultats sur son espace « scratch » de 10Tio, d'accès très rapide. Ces données sont ensuite transférées vers l'espace de travail « datawork », plus volumineux.



Comme le montrent les graphiques ci-dessous, datawork, fin 2018 est alloué à hauteur de 82%, les espaces dédiés aux entités et projets étant prépondérants (67%). Le taux d'occupation global de datawork est de 54%.



L'UMR LOPS du département ODE est destinataire de 78% de l'espace alloué, la partie données satellitaires occupant à elle-seule plus de 1Pio de données.



### 4.3 Statistiques de traitements/calculs

L'infrastructure Datarmor comprend plusieurs « clusters » (ensembles techniques) : des clusters de stockage de données et des clusters de traitement/calcul. Les statistiques d'activité sur les espaces de stockage sont difficiles à faire, car il n'y a pas d'outil ou de métrique partagé pour ce type de ressources. Par contre, les statistiques d'usage de la partie traitement/calcul sont facilement réalisables, au travers du gestionnaire de batch unique utilisé PBSpro.

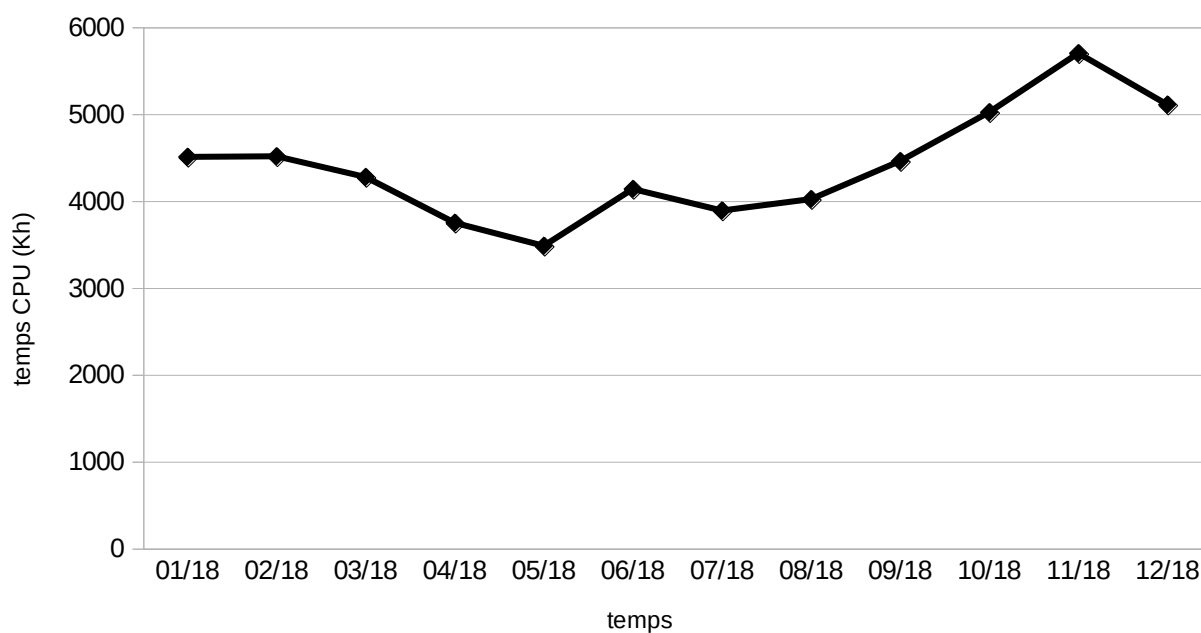
La partie calcul de Datarmor se compose de 4 clusters :

- Le cluster HPCSimple, correspond au cluster de calcul parallèle le plus classique et courant,
- Le cluster HPCHyperThreading, configuré pour utiliser les capacités d'hyperthreading des nœuds de calcul, c'est-à-dire la possibilité de paralléliser 2 calculs élémentaires sur un même cœur de calcul,
- Le cluster Visu de visualisation,
- Le cluster SMP disposant d'une capacité importante de mémoire partagée.

Les configurations techniques des clusters sont détaillées au paragraphe « Configuration technique de Datarmor ».

Un des objectifs de la gouvernance est de proposer le bon cluster de traitement/calcul pour chaque type de code/logiciel à faire tourner ; ces codes/logiciels ne sont pas toujours « optimisés » et peuvent être dépendants de performances en stockage ou en interactif.

### 4.3.1 Temps CPU consommé

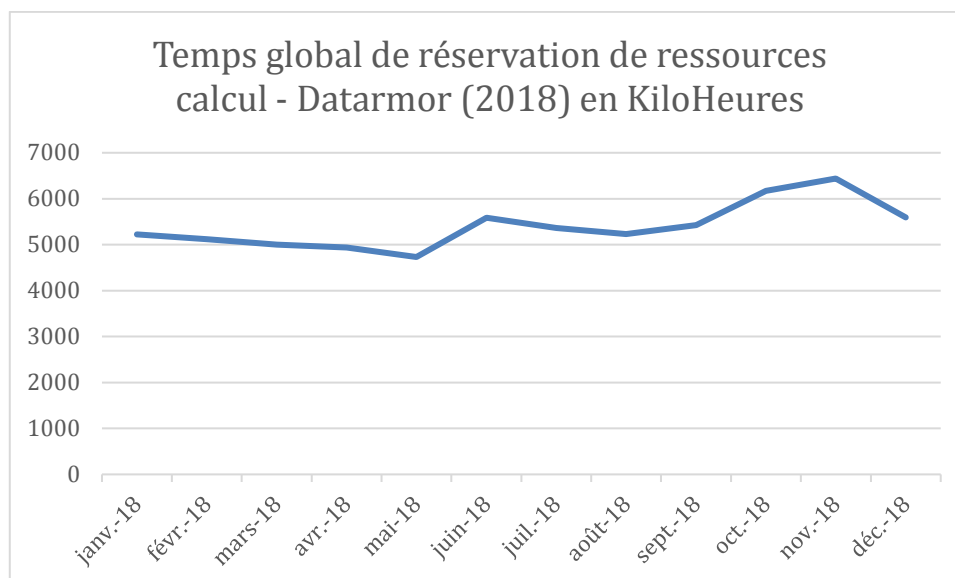


### 4.3.2 Consommation processeur

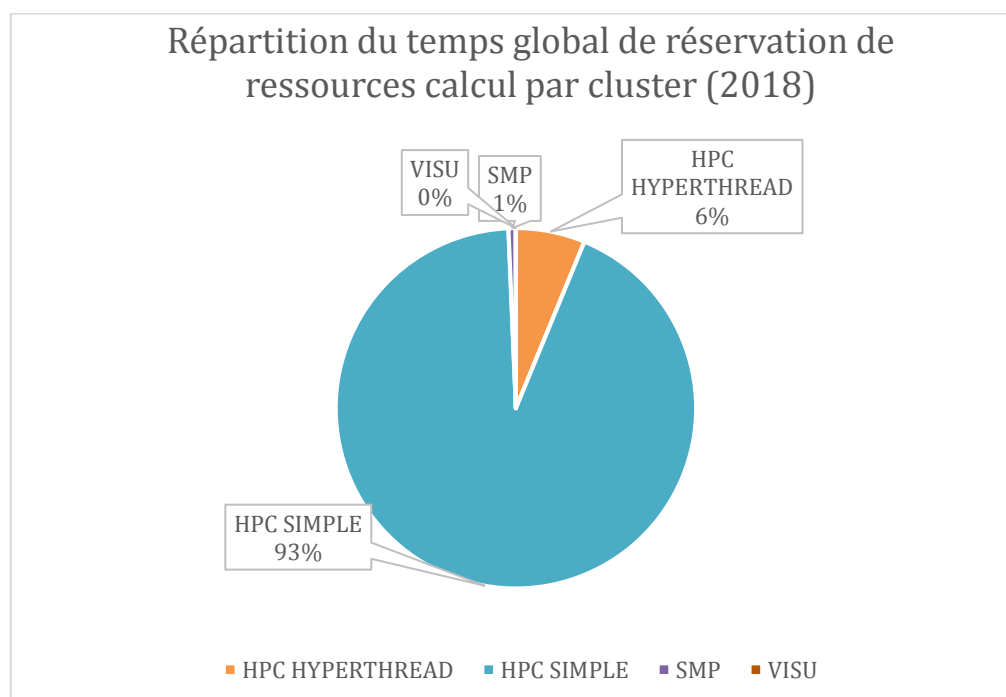
Le temps global de réservation de ressources calcul comprend le temps de calcul effectif mais aussi les temps d'attente induits.

Le temps global de réservation de ressources calcul s'établit entre 5000 et 6000 KiloHeures sur l'année 2018. Il a dépassé les 6000 KiloHeures aux mois d'octobre et novembre.





Le temps de calcul est principalement utilisé sur le cluster HPCSimple.

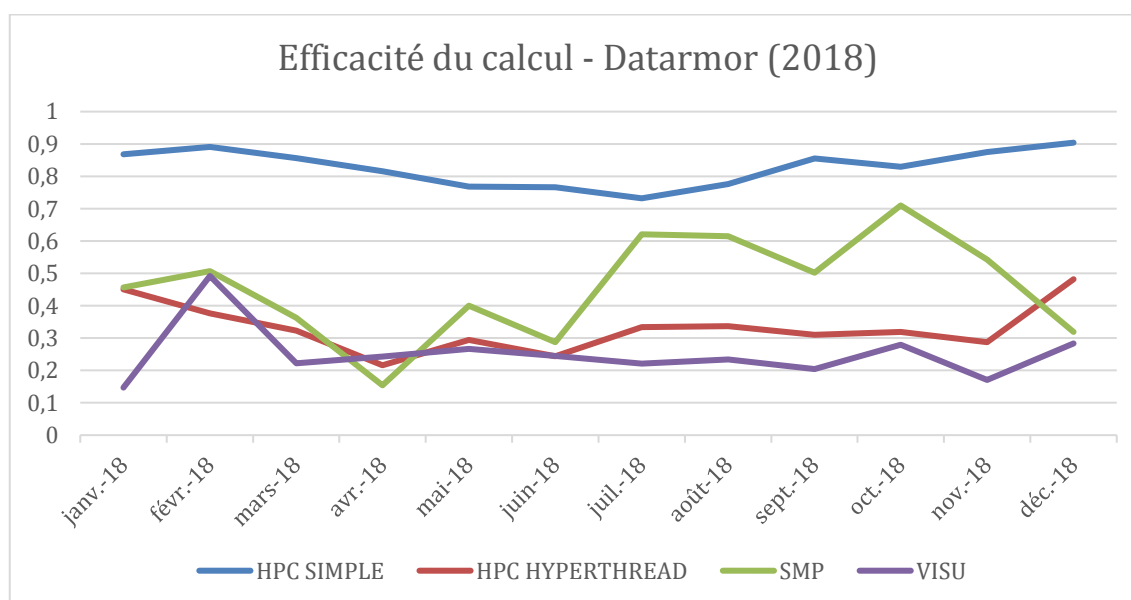
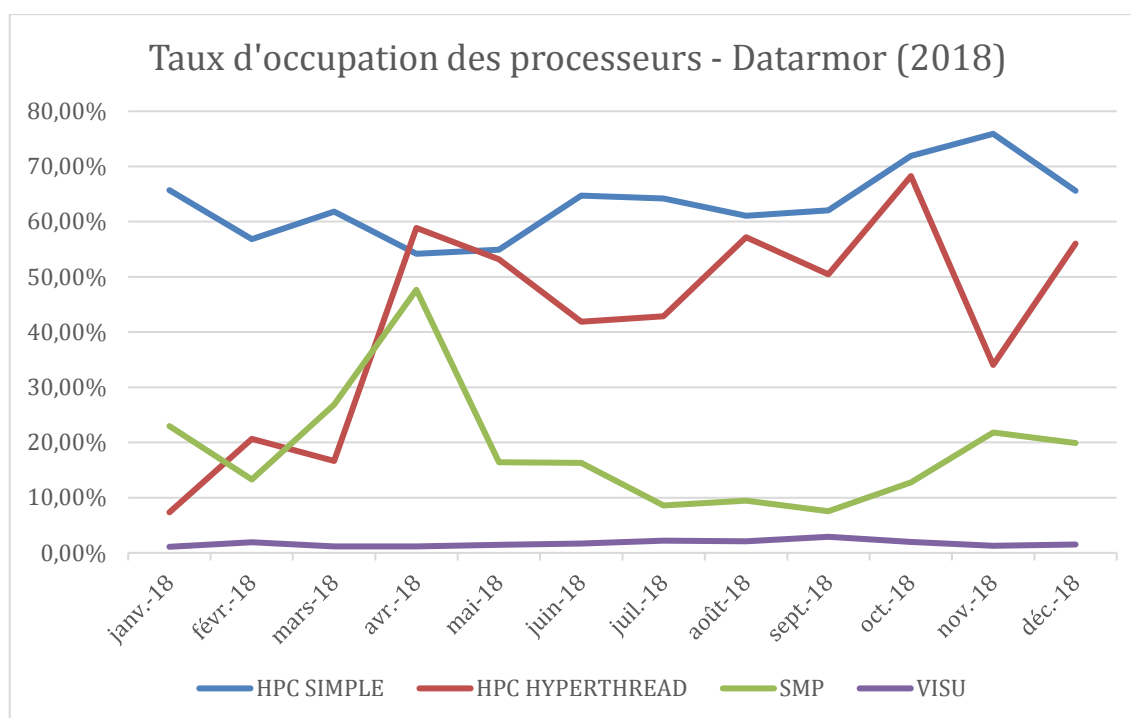


Le taux d'occupation du cluster HPCSimple s'établit autour de 60% avec une pointe en octobre et novembre à 75%.

L'activité sur le HPC Hyperthreading est liée pour une grande partie à des traitements opérationnels, pouvant varier de façon conséquente, elle est en nette augmentation en 2018, ce qui devrait perdurer.

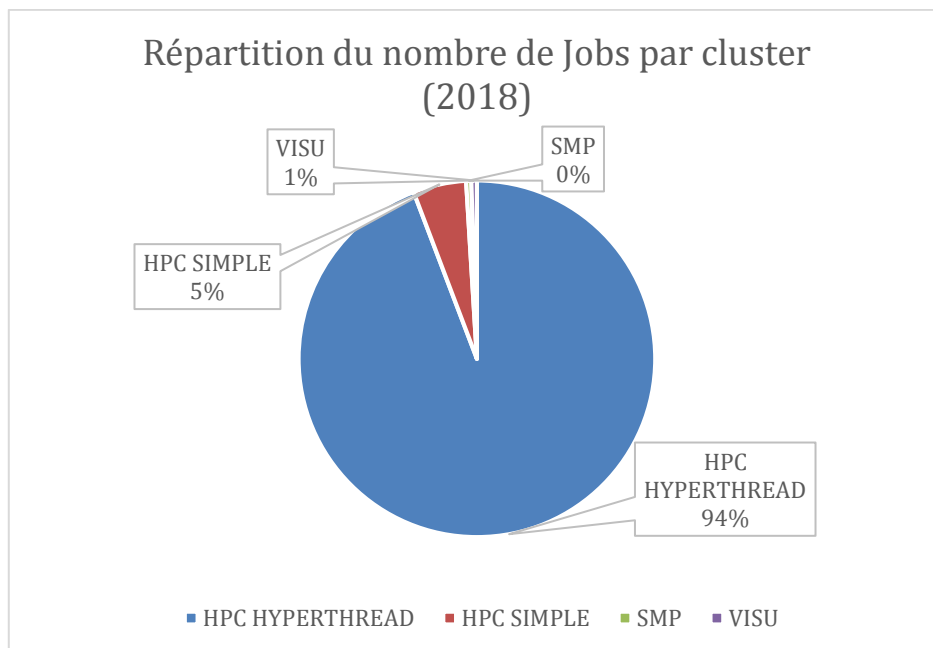
La partie SMP connaît également une progression significative sur 2018.

Le cluster VISU correspond à de nouvelles activités en cours de test.



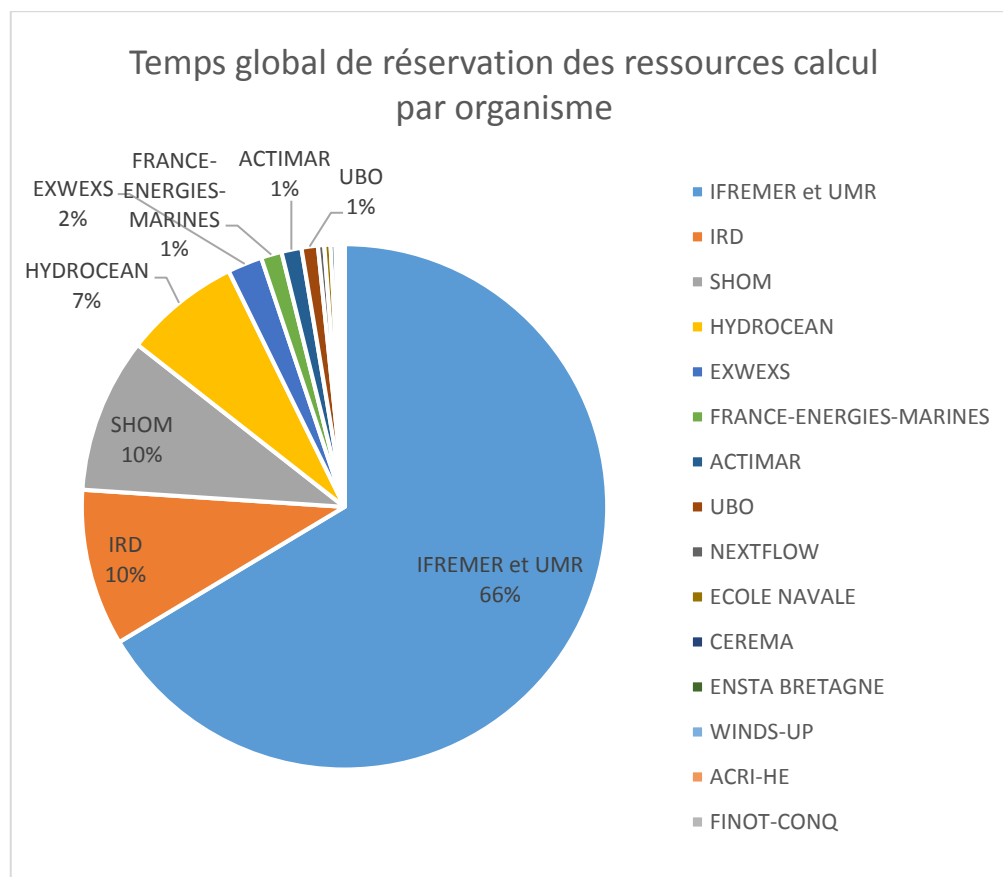
### 4.3.3 Nombre de jobs

Le HPCSimple occupe une très grande part du temps global de réservation des ressources calcul avec des jobs longs, tandis que le HPCHyperthreading est le cluster sur lequel est exécuté le plus grand nombre de jobs, correspondant à de petits calculs.

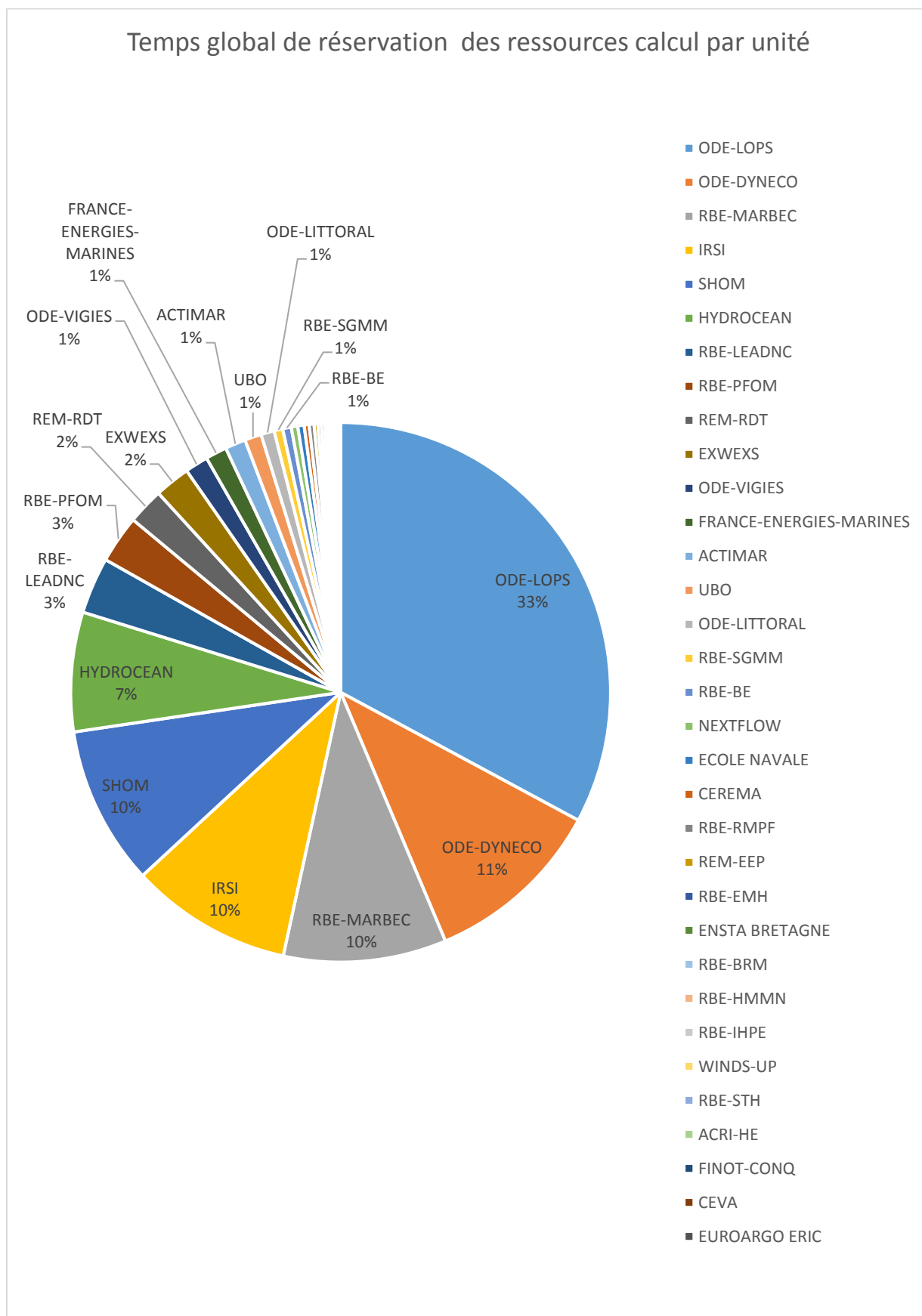


#### 4.3.4 Traitements/calculs par organisme

Les graphiques suivants sont basés sur l'identité des utilisateurs tels que connus dans l'annuaire des comptes informatiques. La catégorie Ifremer comprend non seulement les personnels sous contrat Ifremer mais également les utilisateurs des UMR dont l'Ifremer est partenaire, et les sous-traitants.



## 4.3.5 Statistiques par unité



### 4.3.6 Nombre d'utilisateurs

Le nombre d'utilisateurs ayant utilisé la partie calcul/traitement de Datarmor par l'outil de gestion de traitement batch PBS est de 445 pour 2018.

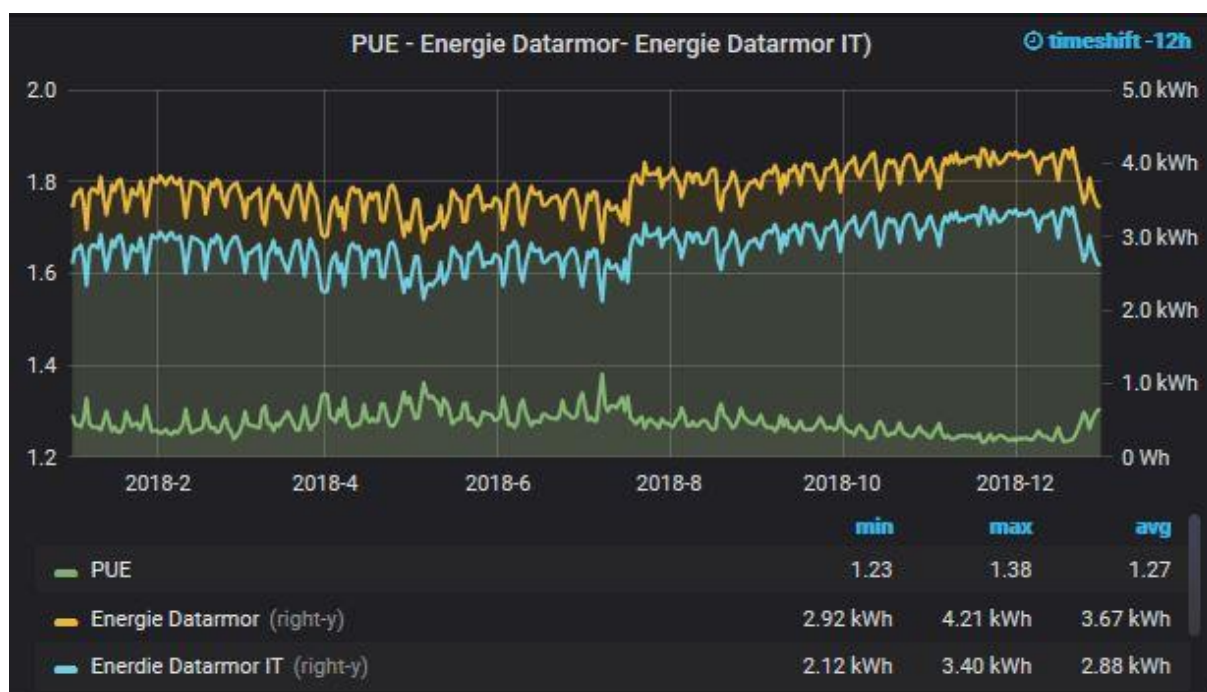
A ces utilisateurs, s'ajoutent ceux qui accèdent aux données stockées dans l'espace Datarmor, par exemple sur Dataref. Cette utilisation, plus diffuse, ne peut se comptabiliser aussi finement que pour le calcul HPC. Potentiellement, l'ensemble des comptes informatiques du réseau Intranet-Ifrermer y a accès, soit environ 4000 utilisateurs.

Certaines données stockées sur l'architecture sont également en diffusion externe publique ou restreinte ce qui élargit leur sphère d'utilisation.

## 4.4 Statistiques logistiques

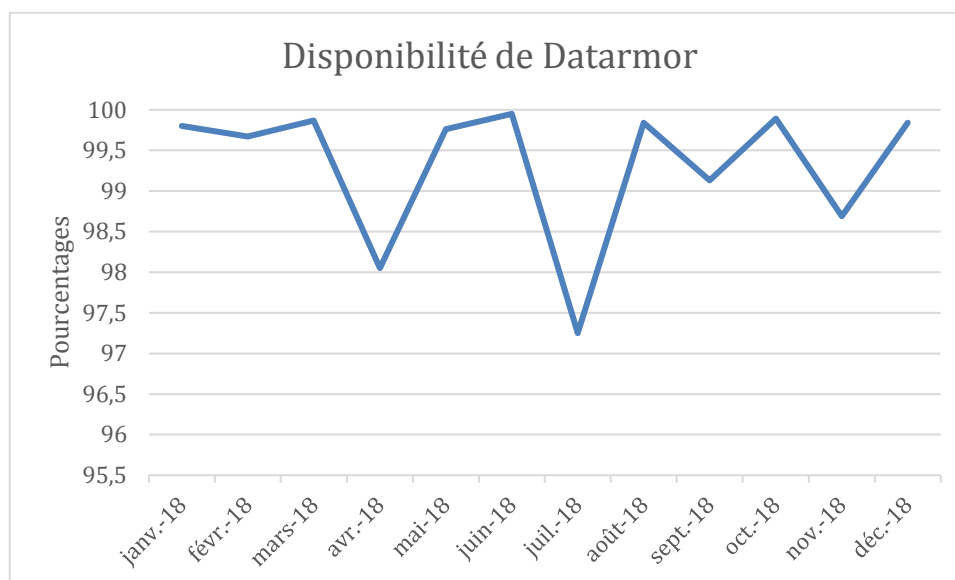
Le PUE (Power Usage Effectiveness) est l'indicateur d'efficacité énergétique utilisé pour un centre d'exploitation informatique. C'est le rapport entre l'énergie totale consommée par la configuration globale (machine informatique + systèmes de refroidissement : compteur Energie Datarmor) et l'énergie consommée par les systèmes informatiques seuls (Compteur Energie Datarmor IT). Le PUE moyen constaté de Datarmor en 2018 sur est de 1,27.

La charge de calcul sur Datarmor a été plus forte au dernier trimestre 2018, et par conséquent, l'énergie consommée plus forte, ce que l'on voit sur les courbes jaune (Energie Datarmor) et bleue (Energie Datarmor IT). On peut voir que pendant cette période de forte activité, le PUE baisse légèrement, l'architecture Datarmor reste efficace énergétiquement à pleine puissance.



## 4.5 Disponibilité

L'indicateur global de disponibilité Datarmor inclut des mesures réseau, espace de stockage et nœuds de calcul pour représenter la disponibilité de l'infrastructure au plus proche de son utilisation. Sur 2018, le taux mensuel le plus bas est le taux de juillet présentant une disponibilité de 97,25%. En juillet, l'espace de stockage des données de travail, datawork, a été augmenté de 1,2Po utiles. Cela a occasionné son indisponibilité pendant quelques heures.



## 5 Services et Ingénierie du PCDM

### 5.1 Upgrade 2018 dans le cadre du CPER

Prévu sous l'appellation de « tranche 2 » du projet CPER Datarmor, l'acquisition et l'installation d'extensions de stockage sur la configuration a été réalisée les 17 et 18 juillet 2018, avec un impact faible sur l'exploitation. L'ajout de 1,3Pio d'espace de stockage DATAWORK a permis en fin 2018 de répondre aux demandes d'espaces exprimées par les équipes scientifiques auprès du GT données.

Cette intervention a aussi permis la mise à jour de quelques firmwares dans le but de surmonter des bugs identifiés.

### 5.2 La bioinformatique

Le domaine de la bio-informatique occupe une place à part dans le paysage de l'informatique pour les sciences marines, en raison de la spécificité des données manipulées (très majoritairement des séquences d'ADN & ARN) et des logiciels utilisés. Les travaux réalisés pour la bio-informatique peuvent se diviser en 3 catégories :

1. la mise à disposition de logiciels « utilisateurs » complètement dédiés à cette discipline,
2. l'exploitation de ces logiciels sur les données des équipes de recherche,
3. la bancarisation et la sécurisation des données de bioinformatiques.

#### 5.2.1 La mise à disposition des logiciels

Le Service de Bioinformatique de l'IFREMER met à disposition une logithèque de plus de 160 outils couvrant un large panel d'applications : metabarcoding, metagénomique, SNP calling, assemblage de génomes, mapping de reads, génotypage, RNA expression, génomique des populations, analyse d'arbres phylogénétiques, prédiction de gènes, profilage de communauté, analyse de méthylation, etc.

Les outils sont installés sur DATARMOR indépendamment les uns des autres en utilisant deux technologies de conteneur logiciel : Conda et Docker. Ils peuvent être exécutés soit en ligne de commande soit via un portail web Galaxy. Via la ligne de commande, les logiciels peuvent être lancés de manière autonome ou combinés dans des pipelines à façon au moyen de Snakemake<sup>1</sup>.

---

<sup>1</sup> <https://snakemake.readthedocs.io/en/stable/>

Les utilisateurs disposent d'un catalogue de tous ces logiciels qui leur permet d'avoir toujours accès à :

- la liste des outils disponibles
- le mode d'utilisation (ligne de commande vs. Galaxy)
- la classification EDAM<sup>2</sup>
- la documentation
- Ce catalogue est géré au moyen d'un logiciel libre conçu par le Service Bioinformatique, ToolDirectory, disponible en ligne<sup>3</sup>.

## 5.2.2 Les applications de bio-informatique

### - **Analyse ADN environnemental**

L'une des activités principales de la bioinformatique à l'IFREMER concerne l'analyse de données issu de l'ADN environnemental. Ce dernier provient de prélèvements d'échantillons réalisés en mer, le long des côtes, à la surface des océans voir dans les profondeurs (sédiments). Ces études permettent de comprendre la diversité des espèces vivantes dans un lieu, non pas via de la biologie d'observation, mais via de nouveaux outils numériques exploitant l'ADN présent dans ces milieux.

Plusieurs logiciels sont nécessaires pour réaliser ces analyses étape par étape (nettoyage des séquences d'ADN, clustering, attribution taxonomique, analyse statistique). Ces différentes parties du calcul sont réalisées au moyen de pipelines coordonnés par le système Snakemake. L'équipe du Service de Bioinformatique de l'IFREMER a mis en place ce type de pipeline qui est appliqué, notamment, sur l'analyse des données des projets ABYSS (analyse de la diversité écologique des grands fonds ; travail en collaboration avec l'équipe MARBEC) et PALMIRA (analyse paléogénomique en Rade de Brest ; travail en collaboration avec l'équipe DYNECO).

### - **Annotation fonctionnelle de génomes**

L'annotation fonctionnelle d'un génome consiste à déterminer la fonction de chaque gène identifié sur le patrimoine génétique (le génome) d'un organisme, quelle que soit sa nature (un virus, une bactérie, une algue, ...). Ce travail se fait, notamment, au moyen de logiciels capables de comparer des séquences d'ADN nouvellement identifiées avec des répertoires de séquences connues. Le challenge consiste à réaliser ces opérations de calcul dans des conditions de big data où plusieurs To de données sont traitées.

L'équipe du Service de Bioinformatique de l'IFREMER a mis en place un pipeline exécutant les logiciels de comparaison BLAST et InterProScan de manière optimale sur DATARMOR. Il a ainsi été possible de réduire des temps de calcul estimés à plusieurs semaines, en utilisant un unique PC, à quelques heures, en distribuant les comparaisons des banques de données. Les résultats obtenus sont alors dépouillés par les biologistes sur une simple station de travail au moyen du logiciel graphique Blast2GO.

<sup>2</sup> <http://edamontology.org/page>

<sup>3</sup> <https://github.com/ifremer-bioinformatics/ToolDirectory>



Le pipeline repose sur l'utilisation des BeeDeeM-Tools, une série d'utilitaires disponibles en logiciel libre<sup>4</sup>.

### 5.2.3 La bancarisation des données de bioinformatique

Le Service de Bioinformatique de l'IFREMER, en lien avec le SISMER, a en charge la mise en place d'un plan de gestion des données de bioinformatique issues des équipes de recherche de l'institut.

Cette gestion implique différentes étapes, notamment :

1. Le stockage physique des données sur l'espace de données DATAREF associé à DATARMOR ; les données sont ainsi placées à proximité du calculateur ;
2. Le classement des données au moyen de l'ontologie de bioinformatique EDAM<sup>5</sup> ;
3. Le positionnement des droits d'accès et d'utilisation des données ;
4. Le catalogage des métadonnées décrivant un projet et sa mise en accès via Sextant<sup>6</sup>
5. L'archivage des données sur bandes magnétiques (sécurisation à long terme)

## 5.3 Organisation des Web services

### 5.3.1 Virtualisation d'applications Web sur la grappe ClusterWeb

L'utilisation de machines physiques avec InfiniBand sur la grappe "Dataweb" apportent son lots d'avantages et d'inconvénients.

Indéniablement, les performances réseaux et d'accès aux disques de données datawork/dataref seront meilleures que via une interface 10 Gbps. Cependant, la présence d'une carte InfiniBand:

- empêche le déplacement automatique des VM « VMware » en cas de crash matériel ; conduisant à une interruption de service le temps qu'un administrateur RIC intervienne
- nécessite des licences et l'expertise de HPE pour installer les drivers GPFS / Lustre dans les VM ; une erreur de configuration pouvant conduire à la compromission des espaces datawork / dataref.

Ces problématiques ont conduit RIC à mener une pré-étude sur l'utilisation des conteneurs Docker en production et notamment sur l'orchestrateur de conteneurs Kubernetes.

La conteneurisation permettrait dans le cas des machines dataweb avec InfiniBand :

<sup>4</sup> <https://gitlab.ifremer.fr/bioinfo/BeeDeeM-Tools>

<sup>5</sup> <http://edamontology.org/page>

<sup>6</sup> <https://sextant.ifremer.fr>

- d'éviter l'installation de nouveaux drivers GPFS / Lustre et l'achat de licences ; les conteneurs réutilisant les points de montage définis sur la machine hôte
- de maîtriser l'impact des applicatifs installés sur les machines dataweb ; les applicatifs dans les conteneurs n'ayant aucun accès par défaut au système de fichiers de l'hôte
- de maîtriser les ressources (CPU, RAM) utilisées par chaque conteneur

L'orchestrateur Kubernetes apporte cette maîtrise des ressources et répartit également les conteneurs sur une grappe de serveurs en fonction de la charge des machines à un instant T et de critères définis par l'administrateur (ex : présence obligatoire de GPU ou de disques SSD).

L'utilisation des conteneurs et d'un orchestrateur offre également des perspectives quant à la gestion de l'infrastructure historique de RIC (ex : cloisonnement de toutes les applications web dans un conteneur, contrôle plus fin des ressources utilisées par une application web, cloisonnement d'un « attaquant intrus » en cas de compromission d'une application web, ...).

### 5.3.2 Plate-forme web JupyterHub

L'amplification, sur 2018, de la demande pour l'utilisation de notebooks Jupyter, des interfaces de programmation interactive, a conduit à l'installation sur Datarmor de la plate-forme <https://datarmor-jupyterhub.ifremer.fr>, qui donne l'opportunité à tous les utilisateurs d'ouvrir, depuis un navigateur, des notebooks utilisant les ressources du calculateur, et ce, de manière standardisée.

Ces notebooks sont utilisés autant à des fins de test/développement que pour de la fouille/exploration interactive de données, ainsi que dans le cadre de formations universitaires.

### 5.3.3 Plate-forme Elastic Search

Elasticsearch est un moteur de recherche NoSQL qui offre une complémentarité aux SGBD de par ses fonctionnalités de recherche (recherche full-text, recherche en fonction d'une localisation spatiale...).

Elasticsearch est communément associé à un autre service : Kibana. Ce service facilite la création de widgets graphiques tout en minimisant les coûts de développement. Ces widgets sont par ailleurs interactifs ; offrant ainsi de nouvelles opportunités pour les fouilles de données.

Dans le cadre d'une installation mutualisée, plusieurs problématiques sont soulevées : performances, gestion des droits et cloisonnement des index par utilisateurs/groupes d'utilisateurs, etc...

Concernant les performances, des benchmarks ont été réalisées pour déterminer l'architecture cible principalement pour le stockage (stockage sur la baie GPFS, stockage sur l'espace cache web...). Du fait de la distribution des serveurs Elasticsearch et des nombreux

I/O disques, ce type de technologie requiert des disques locaux et non pas des baies de disques. Ceci a conduit à l'acquisition d'une nouvelle plate-forme matérielle :

- 2 machines hot-tiers, équipées de disques SSD, pour héberger les index régulièrement mis à jour: dataelastic-hot[12]
- 2 machines warm-tiers, équipées de disques rotatifs, pour héberger les index ne nécessitant pas ou peu de mise à jour : dataelastic-warm[12]

Concernant la gestion des droits et des utilisateurs, Elasticsearch propose une solution payante hors de prix : X-Pack. De fait, la solution SearchGuard a été choisie comme meilleure alternative en terme de prix, fonctionnalités et pérenité de la société réalisant ce produit.

La plateforme Elasticsearch sécurisée, et le service Kibana associé, seront opérationnels au 1er trimestre 2019.

Les premiers services demandeurs sont notamment :

- les sites Archimer, Seanoe (dépôts de publications ou de jeux de données, contre obtention d'un DOI)
- les sites de la flotte océanographique
- Sextant (Infrastructure de Données Géographiques)
- l'application Naiad pour le CERSAT (corrélations spatiales et temporelles de données satellite)

## 5.4 Communication/Informations

### 5.4.1 Journée des utilisateurs 2018

La journée 2018 des utilisateurs du PCDM (Pôle de Calcul et de Données pour le Mer) s'est tenue le 17 mai et a rassemblé 80 personnes autour de présentations des premiers résultats obtenus sur la machine Datarmor. Plusieurs sujets avaient trait aux nouvelles organisations de données volumineuses en opération ou en cours de mise au point, illustrant le caractère stratégique du choix des formats et de l'urbanisation des données pour faire face à l'enjeu de la croissance des volumes.

Après les introductions des responsables des instituts partenaires, 2 sessions de présentations ont permis d'alimenter les discussions :

- Datarmor : nouveau périmètre, nouvelles perspectives
- Nouveaux champs d'applications scientifiques,

L'ensemble des présentations est disponible sur <https://wwz.ifremer.fr/pcdm/L-animation/Journee-2018>.

## 5.4.2 Documentation / Site Web PCDM

Le site Web Internet du « Pôle de Calcul et de données pour la Mer » a été actualisé à l'occasion du lancement de Datarmor (<http://wwz.ifremer.fr/pcdm>). Il contient des informations générales sur l'infrastructure Datarmor (configuration technique, gouvernance) et propose un lien vers un « espace utilisateur » (Figure 5-1), qui contient les documentations, modes d'emploi de datarmor, diverses informations d'organisation et de versions logicielles. L'onglet « les données » contient le catalogue des données de référence (Figure 5-2), avec ses métadonnées descriptives au format ISO19115.

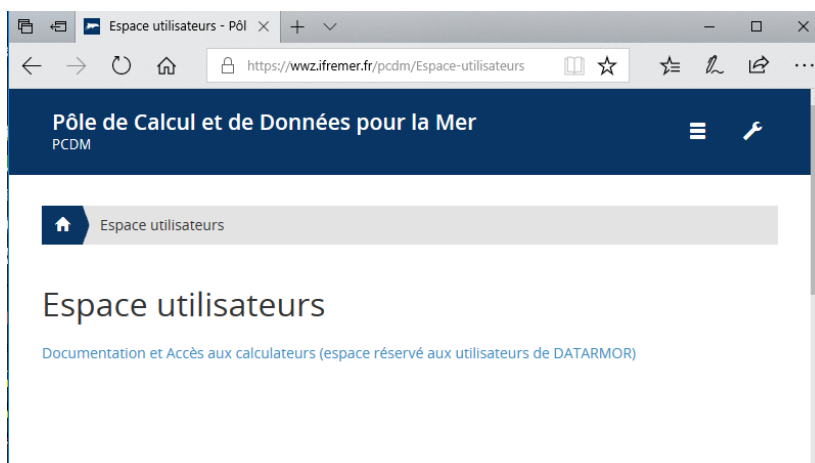


Figure 5-1 'Espace utilisateurs' sur le site web PCDM

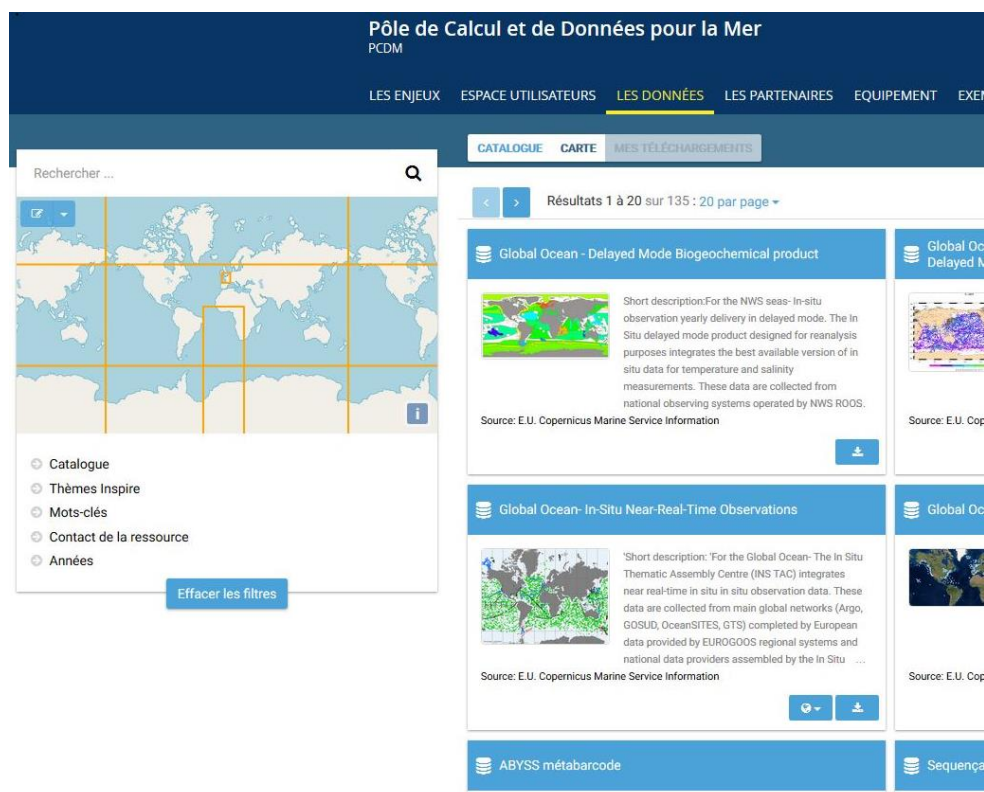


Figure 5-2 'Les données' sur le site web PCDM

### 5.4.3 Enquête utilisateurs

A la demande du COPIL, une enquête « utilisateurs » a été organisée fin 2018/début 2019 pour disposer d'information sur leurs usages et leur satisfaction.

#### 5.4.3.1 Liste des questions du formulaire

NB : le compte utilisateur sert aussi à renseigner le formulaire, et permet d'identifier l'utilisateur, si bien que les questions n'incluent aucun item d'identité ou d'appartenance à une équipe formalisée déjà connue.

<p>1. Votre profil</p> <p>1.1. Métier</p> <p><input type="checkbox"/> Chercheur</p> <p><input type="checkbox"/> Ingénieur / technicien</p> <p><input type="checkbox"/> Doctorant</p> <p><input type="checkbox"/> Post-Doctorant</p> <p>1.2. Thématique scientifique (on peut cocher plusieurs cases)</p> <p><input type="checkbox"/> Bio-informatique</p> <p><input type="checkbox"/> Océanographie physique</p> <p><input type="checkbox"/> Environnement littoral</p> <p><input type="checkbox"/> Géosciences</p> <p><input type="checkbox"/> Halieutique</p> <p><input type="checkbox"/> Mécanique des fluides</p> <p><input type="checkbox"/> Data management/data science</p> <p><input type="checkbox"/> Biologie/Biodiversité</p> <p><input type="checkbox"/> Autre</p> <p>2. Votre niveau d'information</p> <p>2.1. Connaissez-vous ?</p> <p><input type="checkbox"/> Site web PCDM (Pôle de Calcul et de Données pour la Mer)</p> <p><input type="checkbox"/> L'adresse mail de l'assistance aux utilisateurs</p> <p><input type="checkbox"/> Journées des utilisateurs Datarmor</p> <p><input type="checkbox"/> Le catalogue des « données de référence » disponibles sur DATAREF</p> <p>2.2. Connaissez-vous ?</p> <p><input type="checkbox"/> Les différents modes de lancement de jobs sur Datarmor ?</p> <p><input type="checkbox"/> Les différents types d'espace de stockage ?</p> <p><input type="checkbox"/> Le nom du correspondant de votre dept/entité au sein de la gouvernance de Datarmor ? (pour le SHOM, l'ENSTA Bretagne, l'IUEM, l'UBO)</p> <p><input type="checkbox"/> Le réseau de fibres INFUSER ?</p> <p>3. Mode d'accès</p> <p>3.1 Mode d'accès</p> <p><input type="checkbox"/> Par SSH, puis en mode commande</p> <p><input type="checkbox"/> Via une interface web spécialisée (Galaxy pour la bio-info, Jupyter, autre...)</p> <p><input type="checkbox"/> Uniquement par NFS/SAMBA (accès aux données depuis votre poste de travail)</p> <p><input type="checkbox"/> Autres</p> <p>3.2 Décrivez vos autres modes d'accès à Datarmor</p> <div style="border: 1px solid black; height: 20px; width: 100%; margin-top: 5px;"></div>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#### 4. Usage traitements/calculs

##### 4.1 Maitrisez-vous les codes source de vos applications ?

- Oui
- Non (usage basé sur des logiciels existants)

##### 4.2 Quels modèles utilisez-vous ?

- Mars3D
- HYCOM
- WaveWatch III
- TELEMAC
- CROCO
- ROMS
- Autre
- Aucun

##### 4.3 Quels autres modèles utilisez-vous?

##### 4.4 Utilisez-vous un logiciel de couplage ?

- Oui
- Non

##### 4.5 Traitements de données

- Vous développez vous-même vos algorithmes de traitement de données
- Vous utilisez des logiciels

##### 4.6 Quels logiciels de traitement de données utilisez-vous?

#### 5. Espaces de stockage

##### 5.1 Usage des espaces de stockage

- Vous utilisez régulièrement des données de référence issues de DATAREF
- Vous utilisez des espaces de stockage partagés (équipe ou projet)
- Je ne sais pas

#### 6. Autres besoins

##### 4.6 Avez-vous une vision de besoins de traitements présents ou à venir nécessitant une évolution de l'architecture Datarmor ? veuillez préciser.

#### 7. Satisfaction

##### 7.1 Êtes-vous satisfait du fonctionnement de la machine Datarmor ?

- Oui
- Non

##### 7.2 Datarmor rend-elle les services qui étaient attendus ?

- Oui
- Non

##### 7.3 Datarmor a-t-elle ouvert de nouvelles perspectives ?

- Oui
- Non

7.4 L'équipe d'exploitation assure-t-elle les services attendus ?

Oui

Non

7.5 La gouvernance de Datarmor arbitre-t-elle correctement les différentes demandes ?

Oui

Non

7.6 Êtes-vous correctement informé des événements, évolutions, changements réalisés sur Datarmor ?

Oui

Non

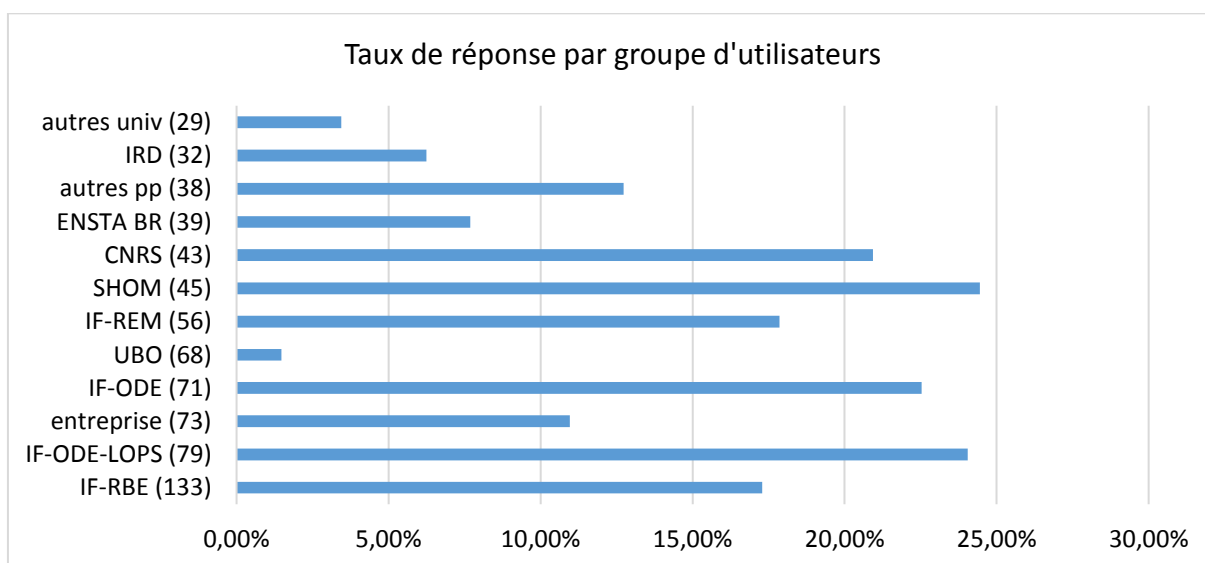
8. Expression libre

8.1 Un commentaire, une remarque ...

### 5.4.3.2 Représentativité des réponses

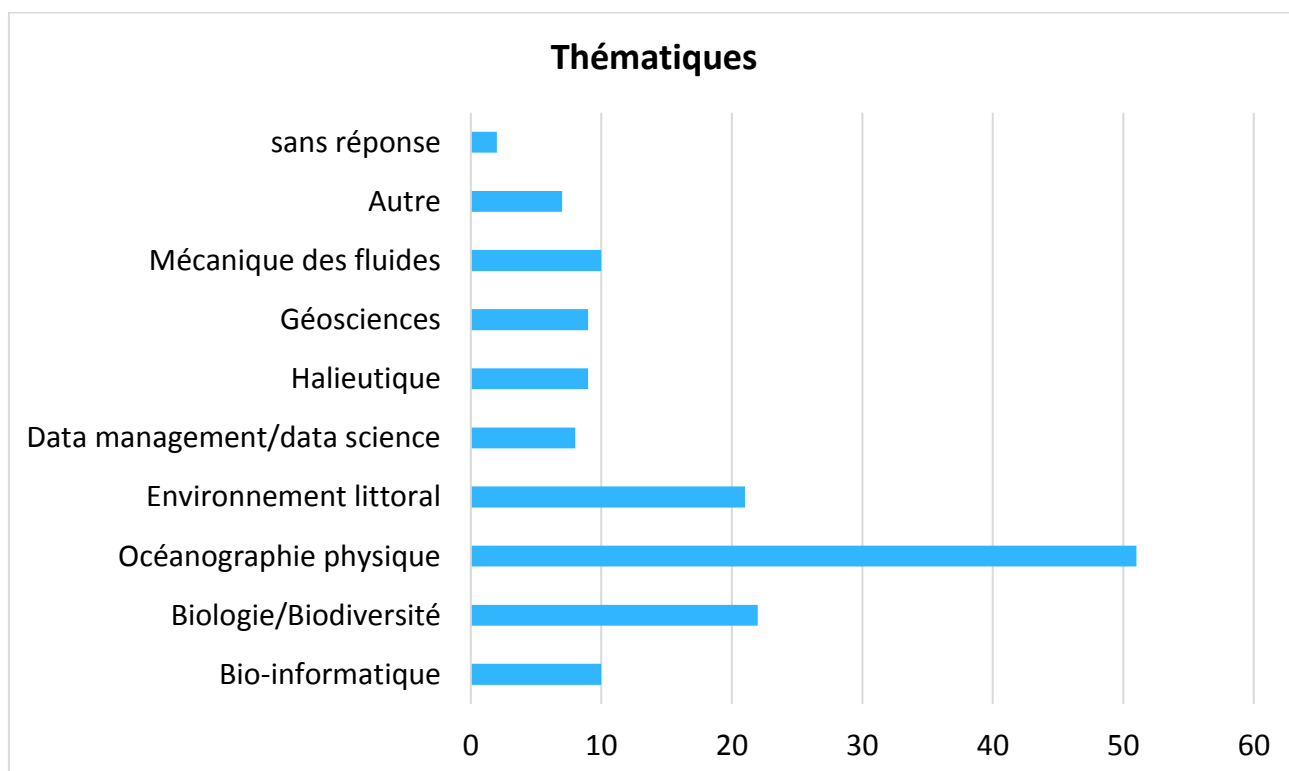
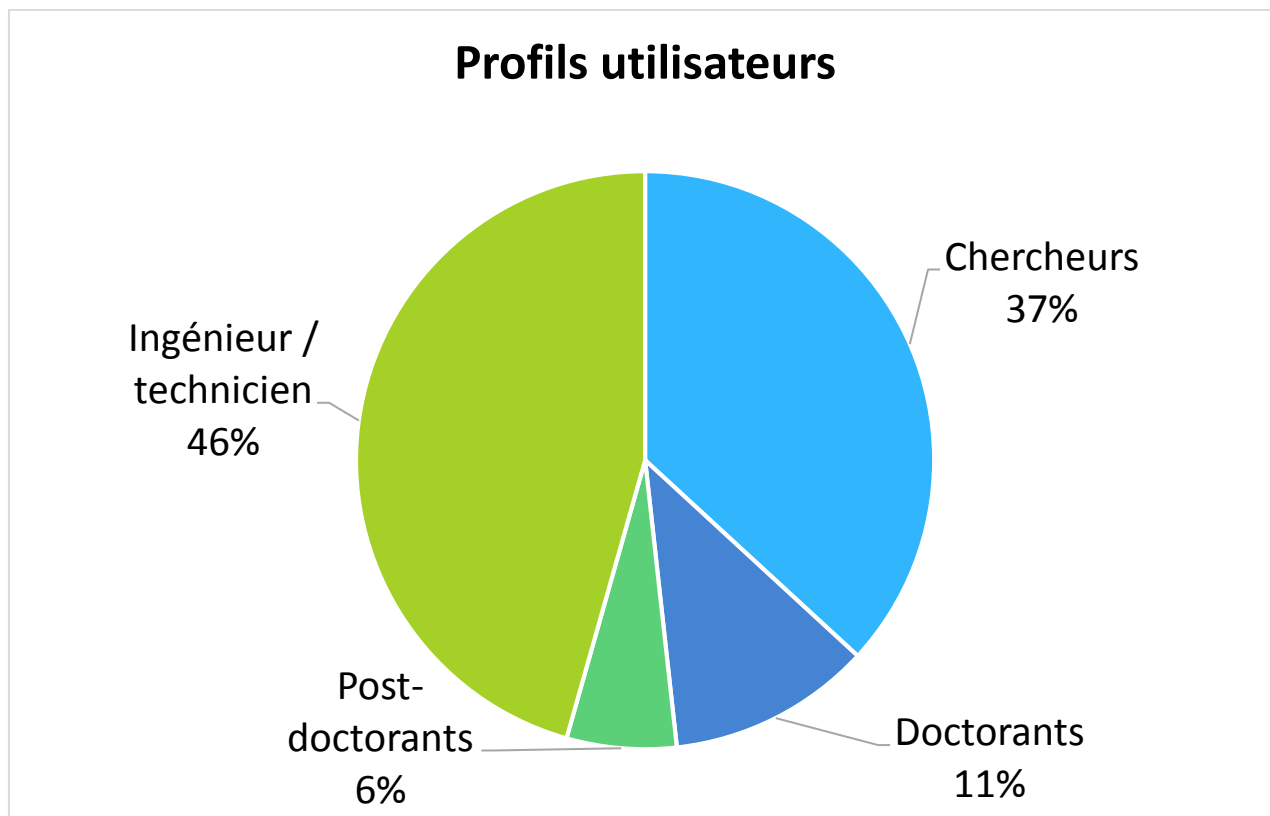
Sur les 752 utilisateurs différents de Datarmor observés en 2018, 114 ont répondu, soit 15,16%.

Le graphe suivant indique les différents taux de réponse par grande entité/unité utilisatrice.



### 5.4.3.3 Profil des répondants

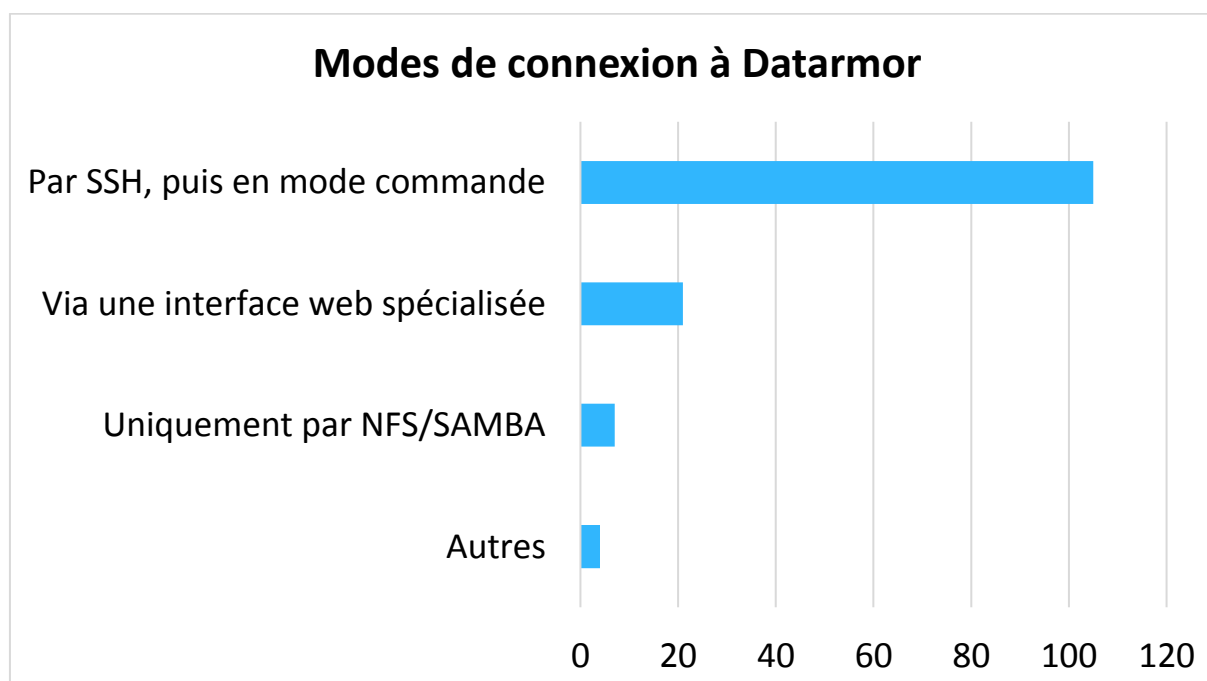
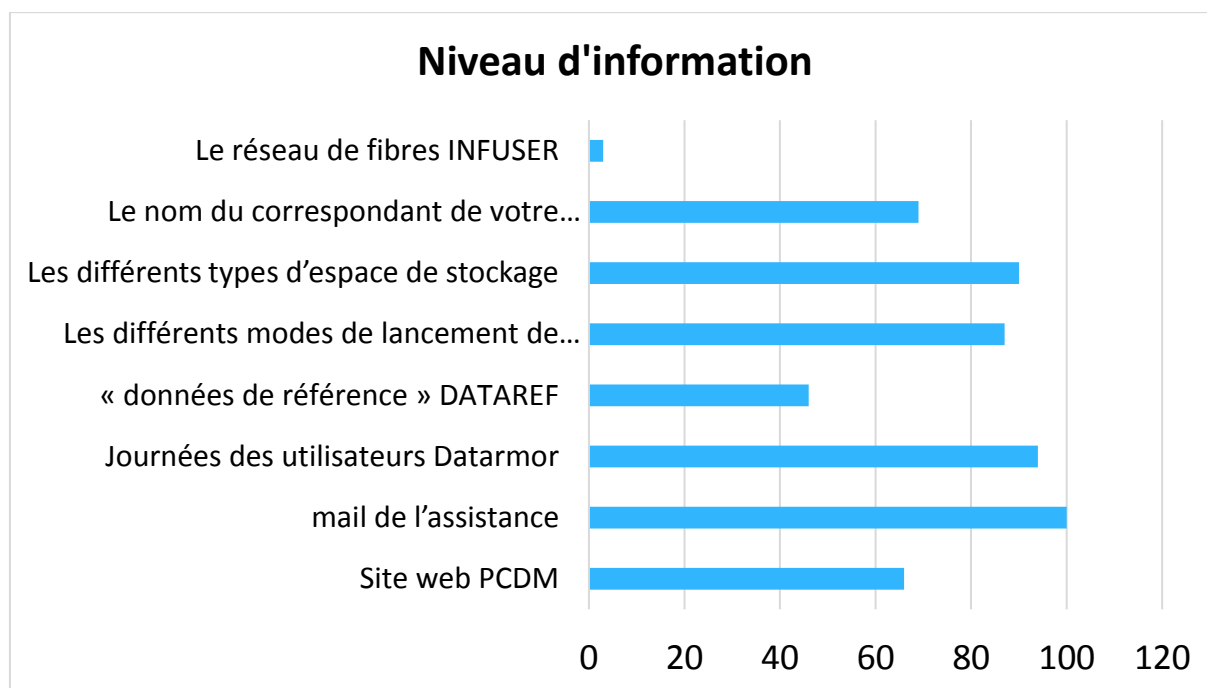
Il s'agit ici de qualifier la population d'utilisateurs : par profil (graphe 1) et par sous-thématique (graphe 2).





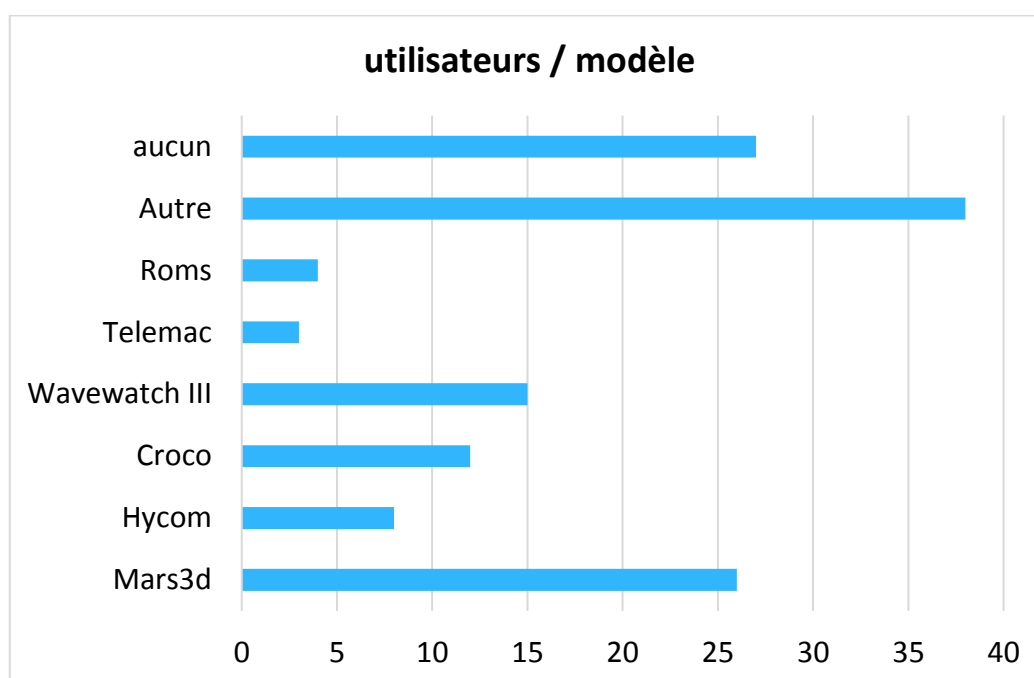
#### 5.4.3.4 Niveau d'information des utilisateurs

Ce lot de questions avait pour but de mesurer le degré d'information des utilisateurs sur le plan technique, mais aussi pour la gouvernance et les différents services disponibles.



### 5.4.3.5 Sur l'usage de modèles

Plus des 2/3 des répondants affirment maîtriser les codes sources de leurs modèles ; environ ¼ d'entre eux utilisent des logiciels de couplage, qui permettent d'associer 2 modèles de 2 domaines différents mais interdépendants (vagues-courants, par exemple). Le graphe suivant montre les distributions des principaux modèles océanographiques connus.

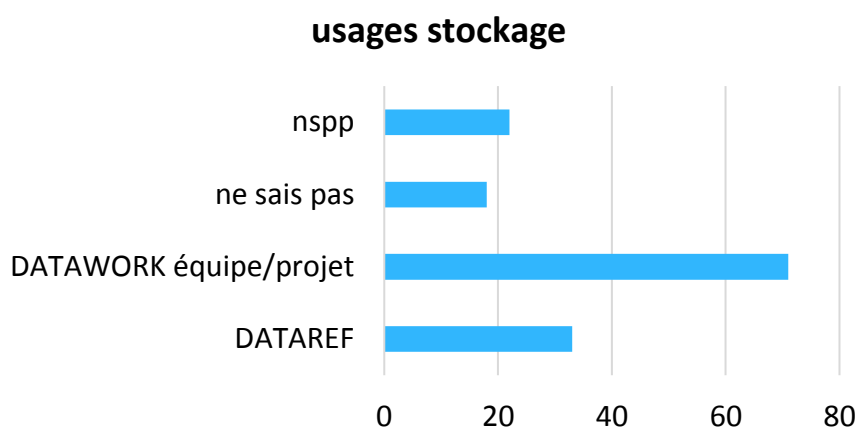


Le texte libre qui accompagnait a permis d'identifier d'autres modèles non cités dans le formulaire de l'enquête :

- MESONH : 5 fois
- NEMO : 5
- WRF : 3
- ECOMARS : 3
- Plus quelques autres

### 5.4.3.6 Outils pour le traitement de données

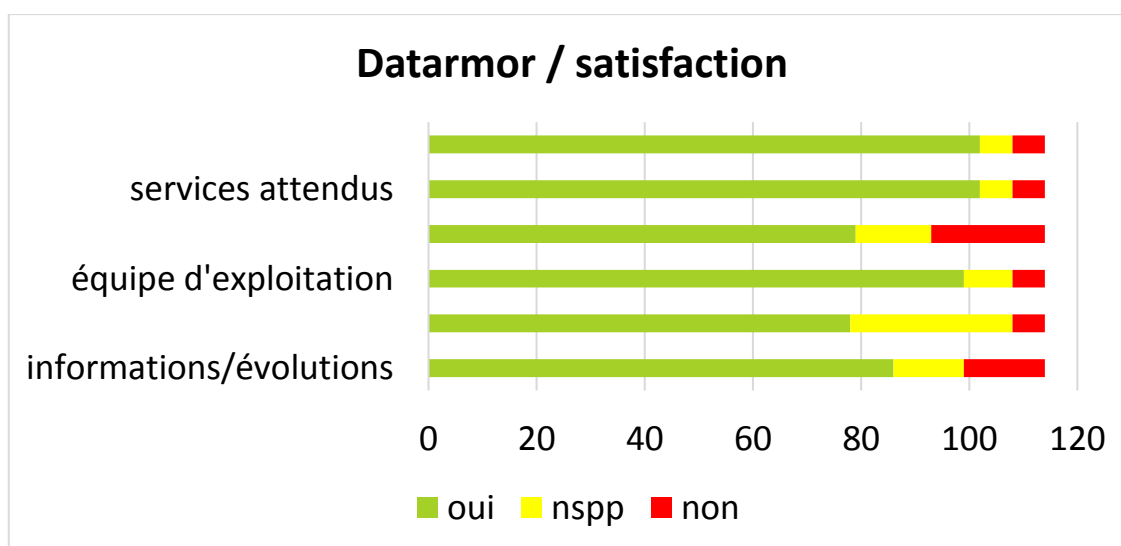
Il semble que le périmètre de ces outils soit interprété de manière variable chez les répondants. Ils utilisent des traitements développés en propre en majorité, et des logiciels « configurables », parmi lesquels Matlab (17 citations dans le texte libre) et « R » (18 citations) soient les plus fréquents. Au niveau des usages des espaces de données, le graphe suivant présente les réponses obtenues.



#### 5.4.3.7 Mesure de la satisfaction

Le graphe suivant est le bilan des réponses au QCM de satisfaction utilisateurs. Plusieurs réponses ont fait état du caractère manichéen du QCM, et que les « oui » et « non » méritaient d'être nuancés. Mais le QCM, c'est aussi une facilité de dépouillement... Les contacts pris directement avec les répondants ont ensuite montré que le périmètre des services faisant l'objet ou non de satisfaction était variable suivant les répondants. Ainsi, certains ont bien jugé de la disponibilité/fiabilité de l'infrastructure et d'autres sur le périmètre des services rendus, avec parfois des lacunes constatées.

Les réponses fournies dans les cases de texte libre ont permis de prendre contact individuellement avec certains répondants, pour préciser des problèmes techniques évoqués.



#### 5.4.3.8 Conclusion de l'enquête

Les réponses à l'enquête sont globalement satisfaisantes et même encourageantes pour l'équipe d'administration technique, mais aussi les personnes du SISMER impliquées dans la gouvernance des données de référence (DATAREF). Il sera sans doute intéressant de refaire d'autres enquêtes de ce type dans le futur, mais les questions gagneront à cibler des sujets plus précis.

## 6 Les exemples d'utilisation du PCDM

### 6.1 Recherche académique

#### 6.1.1 L'exploitation de l'imagerie hyperspectrale pour la caractérisation de la végétation marine : une première sur Datarmor

*Touria BAJJOUK – Ifremer DYNECO/LEBCO*

##### 6.1.1.1 INTRODUCTION

La végétation marine, en particulier les phanérogames marines, est naturellement utilisée comme bio-indicateur de la qualité du milieu. Mettre en place des outils pour suivre son extension et estimer spatialement certains de ses paramètres biologiques permettrait de fournir des informations clé pour aider à la compréhension de sa dynamique spatio-temporelle.

Par ailleurs, l'imagerie hyperspectrale présente un potentiel quant à la caractérisation spatiale des fonds marins et pourrait donc devenir à terme une technique de référence pour la production d'indicateurs d'habitats. Elle rassemble l'information dans de nombreux canaux correspondant chacun à une longueur d'onde du spectre électromagnétique. Chaque pixel est donc un spectre complet de réflectance, permettant une identification précise des matériaux de la scène observée [R1].

L'objectif de la présente étude est d'évaluer les capacités de l'imagerie hyperspectrale pour mieux caractériser la distribution des herbiers de zostères. En effet, les herbiers de zostères peuvent être présents de manière homogène à différentes densités ou en mélange avec d'autres types de végétation marine. Il est donc nécessaire de « démêler » les contributions des types de fonds présents dans chaque pixel de l'image hyperspectrale. Le volume important des images hyperspectrale est jusqu'à présent un frein pour son utilisation dans un cadre opérationnel, l'utilisation du supercalculateur DATARMOR a ainsi été envisagée.

Ce travail se déroule dans le cadre du projet ZOSTERA soutenu par AQUAREF. Il s'appuie sur un jeu de données disponible sur le site du golfe du Morbihan acquis dans le cadre du réseau Rebet Bretagne.

##### 6.1.1.2 ACQUISITION DES DONNEES

Une campagne d'acquisition aéroportée des images hyperspectrales a été réalisée sur le Golfe du Morbihan le 20 août 2013 couvrant une superficie totale de 111,7 km<sup>2</sup>. Le capteur hyperspectral utilisé est le AISA Eagle 1K qui a permis d'enregistrer des informations spectrales sur la plage [400nm ; 970nm], avec une résolution spatiale de 80 cm et une résolution spectrale de 5 nm pour un total de 126 canaux.

Le plan de vol a été élaboré en retenant un chevauchement maximal de 35 %. 38 axes ont été programmés en suivant la progression de la basse mer d'ouest en est pour tenir compte de l'écart de marée important sur ce site (environ 2 heures entre l'est et l'ouest).

### 6.1.1.3 PRINCIPE DE TRAITEMENT

Plusieurs opérations ont été nécessaires préalablement à l'utilisation des images pour les différents traitements développés en vue de l'extraction des informations sur la distribution des herbiers. Il s'agit principalement de corrections atmosphériques, radiométrique et d'orthorectification générant des images en réflectance surface géolocalisées dans le référentiel Lambert 93.

Les images hyperspectrales ont également été corrigées des perturbations du signal dues à la présence de l'eau et éléments présents dans la colonne d'eau. Ce traitement est basé sur le modèle de Lee [R2] et fait appel à des méthodes d'optimisation de problèmes non linéaires

L'application d'un modèle de mélange linéaire combinée à une librairie spectrale des principales composantes du fond, a permis de produire des cartes d'abondance des deux espèces d'herbiers de zostères, *Z. marina* et *Z. noltei* (**Figure 6-1**). Le principe de démelange se base sur une relation linéaire entre les spectres des pixels dans une image hyperspectrale et ceux d'un ensemble de matériaux « purs » considérées comme sources latentes [R1].

### 6.1.1.4 IMPLEMENTATION SOUS DATARMOR

En plus du volume important que représente les images hyperspectrales (254 Go pour le site test et pouvant aller à plusieurs To sur de larges zones), le principe de traitement fait appel à des méthodes d'inversion et d'optimisation itératives très coûteuses en termes de temps de calcul. Un déploiement sur le supercalculateur DATARMOR a donc été nécessaire. Pour l'implémentation des algorithmes, le choix s'est orienté vers les outils libres (le logiciel R), avec une parallélisation des codes (MPI) pour optimiser le temps de traitement. Selon les images, plusieurs nœuds de 28 cœurs chacun et 110 Go de mémoire ont été sollicités pour l'exécution des traitements. C'est en général 224 cœurs et 880 Go de mémoire distribuée qui ont été utilisés pour chaque image. La figure 6-1 illustre le résultat obtenu sur le secteur Est du Golfe du Morbihan.

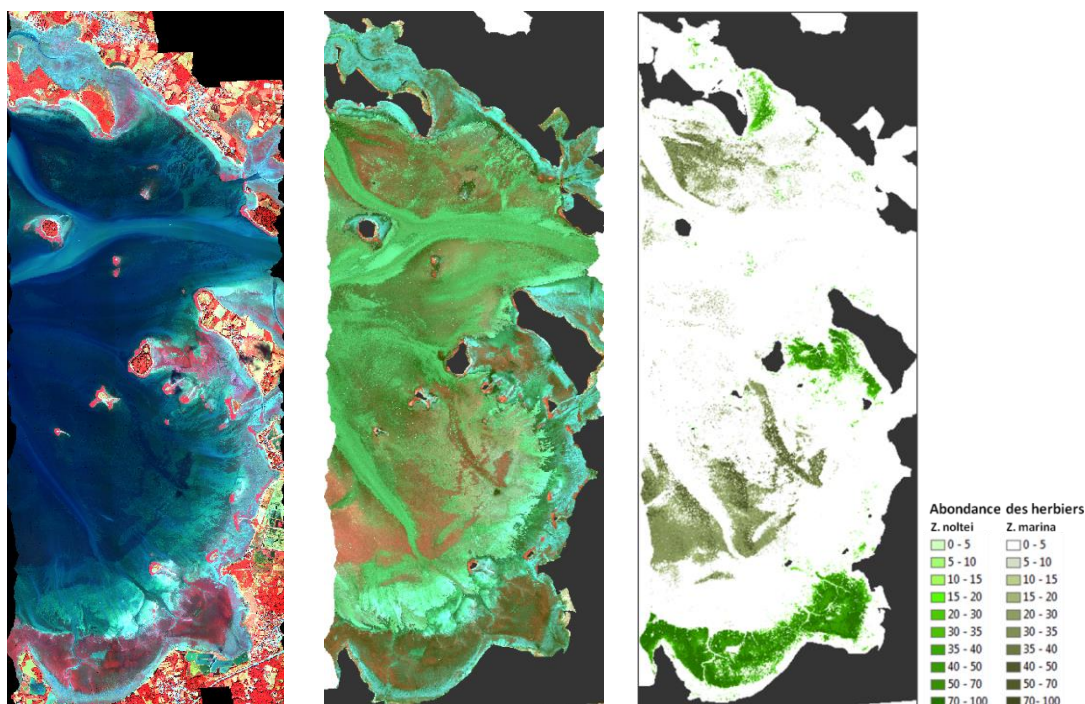


Figure 6-1. : Visualisation RGB (à gauche), image corrigée de l'effet de la colonne d'eau (au milieu) et résultat de cartographie de l'abondance des deux espèces d'herbiers de zostères sur le secteur Est du Golfe du Morbihan (à droite).

Le résultat obtenu est bien représentatif et confirme la détection très performante des herbiers de zostères marines par le capteur hyperspectral, y compris pour les herbiers de zostères marines en zone submergée (précision de détection sur ce secteur de 91 % pour 924 points de validation vidéo).

Ces premiers tests montrent également le potentiel d'utilisation de ce type d'imagerie, qualifiée de « Big Data », en vue de leur exploitation dans un cadre opérationnel tel que la mise en œuvre des directives européennes (DCE, DCSMM et DHFF), ceci grâce aux moyens de calcul et de stockage qu'offrent DATARMOR.

#### 6.1.1.5 REFERENCES

[R1] Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. (2013). *Hyperspectral remote sensing data analysis and future challenges*. *IEEE Geoscience and remote sensing magazine*, 1(2), 6-36.

[R2] Lee, Z., Carder, K.L., Mobley, C.D., Steward, R.G., Patch, J.S., 1999. *Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization*. *Appl. Opt.* 38, 3831–3843. <https://doi.org/10.1364/AO.38.003831>

### 6.1.2 Etude de l'influence des forçages naturels et anthropiques sur la dynamique hydro-sédimentaire en Estuaire de Seine

*Florent Grasso (ODE/DYNECO/DHYSED), Jean-Philippe Lemoine (GIP Seine-Aval), Pierre Le Hir (ODE/DYNECO/DHYSED)*

Le fonctionnement hydro-sédimentaire de l'estuaire de la Seine est principalement soumis aux forçages naturels (vent, vagues, marée, débit liquide). Les différentes conditions hydrométéorologiques ont une influence majeure sur la réponse du système estuarien, en termes d'hydrologie (ex : remontée des eaux marines) et de dynamique des sédiments en suspension (ex : bouchon vaseux, flux résiduels). Néanmoins, les forçages anthropiques (ex : aménagements portuaires, dragages) peuvent également impacter le système hydro-sédimentaire estuarien.

Au cours des 50 dernières années, l'estuaire est passé d'un système quasi-naturel à un système fortement anthropisé marqué par l'approfondissement et l'endiguement des chenaux de navigation, le rétrécissement de l'estuaire à l'embouchure et l'intensification des dragages d'entretien. Des simulations réalisées sur le calculateur DATARMOR (utilisation de 168 processeurs) à partir du modèle hydro-sédimentaire MARS3D-MUSTANG (Ifremer) ont permis de montrer que pour des configurations morphologiques contrastées (1960, 1975 et 2010), l'hydrodynamique, l'hydrologie et la dynamique sédimentaire ont significativement évoluées (projet ANPHYECO, Seine-Aval 5). En particulier, la dynamique du bouchon vaseux était fortement dominée par l'influence du pompage tidal en 1960 et la stratification était peu développée. De manière contrastée en 2010, l'approfondissement et le rétrécissement de l'estuaire ont renforcé la remontée des eaux marines et intensifié la stratification dans l'estuaire, s'illustrant par une dynamique du bouchon vaseux conjointement influencée par le pompage tidal et la circulation barocline.

De plus, le fonctionnement hydro-morpho-sédimentaire d'un estuaire dragué pour les besoins de la navigation peut se révéler très sensible au déplacement des masses sédimentaires liées aux opérations de dragage et au dépôt de matériaux dragués. C'est pourquoi le projet MEANDRES (Seine-Aval 6) a pour objet d'étudier l'effet des transferts anthropiques de sédiments sur le fonctionnement hydro-morpho-sédimentaire de l'estuaire de la Seine. En s'appuyant sur le même modèle hydro-sédimentaire implémenté sur DATARMOR, une technique de marquage numérique des particules permet de réaliser un suivi spatio-temporel des sédiments dragués puis clapés en mer. Cela sert à caractériser la part de ces matériaux susceptibles de réalimenter le bouchon vaseux, les vasières latérales ou les bancs d'embouchure, ainsi que les temps de transferts associés (Figure 6-2). L'influence des stratégies de dragage sur l'évolution à moyen et long terme (20 à 50 ans) de la morphologie de l'estuaire et de la nature de sa couverture sédimentaire est également considérée, selon différents scénarios prospectifs.

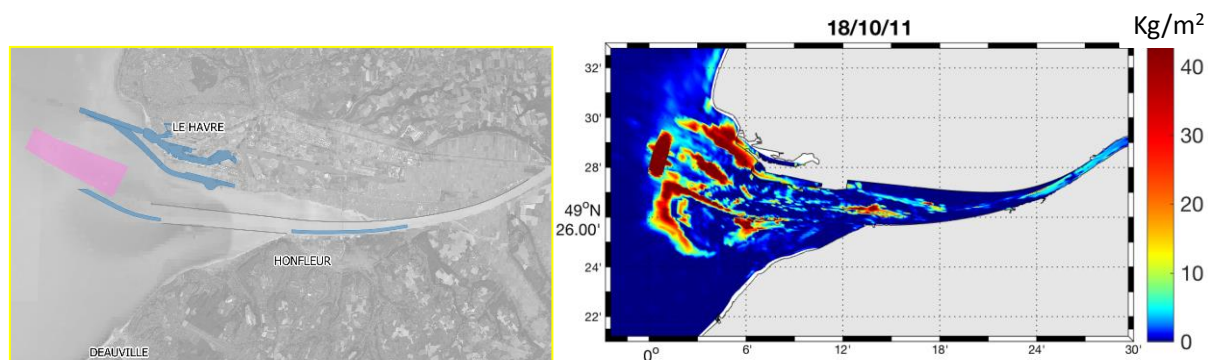


Figure 6-2. Embouchure de l'estuaire de la Seine : (a) zones de dragages (bleu) et de clapage du Kannik (rose) ; (b) dispersion des sédiments clapés sur le site du Kannik après 3 ans de simulations.

### 6.1.3 Influence des tourbillons sur la variabilité intrinsèque basse fréquence et grande échelle de l'Atlantique Nord

*Antoine Hochet (UBO/LOPS); Thierry Huck (CNRS/LOPS); Florian Sévellec (CNRS/LOPS); Alain Colin de Verdière (UBO/LOPS); Olivier Arzel (UBO/LOPS); Matthew Mazloff (SCRIPPS<sup>7</sup>); Bruce Cornuelle (SCRIPPS)*

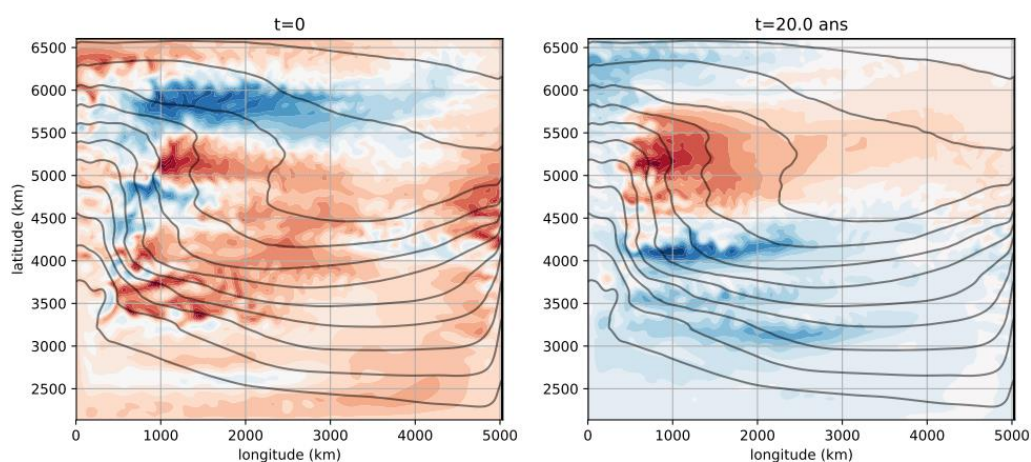


Figure 6-3: Anomalie de température de surface de l'océan (rouge positive, bleu négative) à 20 ans d'intervalle dans un bassin idéalisé de type Atlantique NORD. Superposées en noir, les lignes d'isotherme moyenne. Résolution: 10 km.

Des observations montrent que la température de surface moyennée spatialement sur l'Atlantique Nord présente des variations temporelles sur des périodes de 20 à 60 ans. Ces variations océaniques sont corrélées avec des variations climatiques: précipitations sur le Sahel, sur l'Europe et l'Amérique du Nord, activité cyclonique en Atlantique Nord.

<sup>7</sup> Scripps Institution of oceanography San Diego USA



De nombreuses théories tentent de rationaliser ce phénomène. Certaines suggèrent une dynamique impliquant un couplage entre l'océan et l'atmosphère, d'autres impliquant uniquement l'océan. C'est à cette dernière catégorie qu'appartient cette étude. Des travaux précédents ont montré que cette variabilité émerge spontanément d'un océan mis en mouvement par des flux de chaleur à sa surface, mais négligent, pour la plupart, l'effet des tourbillons qui apparaissent lorsque la résolution spatiale augmente. Deux hypothèses peuvent être faites sur l'effet de ces tourbillons. La première est qu'en créant du mélange à petite échelle et haute fréquence, ils s'opposent à l'existence de l'anomalie grande échelle. La deuxième, soutenue par un ensemble d'études récentes sur la cascade inverse temporelle d'énergie cinétique, postule qu'à l'inverse, les tourbillons ont la capacité d'exciter la variabilité basse fréquence. Des résultats préliminaires tendent à privilégier la première hypothèse.

Les capacités de Datarmor sont ici mises à profit pour simuler l'océan à haute résolution (10 km) sur des périodes de temps longues (au moins 500 ans) permettant d'étudier la coexistence des tourbillons et de l'anomalie basse fréquence. Pour cela, le modèle MITgcm est utilisé dans une configuration idéalisée de l'Atlantique Nord et plusieurs simulations permettent d'étudier la sensibilité des résultats à différents paramètres du modèle. Sur la figure 6-3, deux cartes de température de surface, obtenues à 20 ans d'intervalle, illustrent la cohabitation d'une structure de grande échelle et basse fréquence (d'une période d'environ 40 ans) avec des tourbillons.

Il s'agit d'un projet financé par une bourse Marie Curie (MSCA-IFGF) en collaboration avec le Scripps Institute of Oceanography à San Diego.

#### 6.1.4 Modélisation haute résolution de la convection profonde en mer Méditerranée Nord-Occidentale

*Martial Boutet (LOPS, Ifremer), Pierre Garreau (LOPS, Ifremer), Franck Dumas (Shom)*

##### 6.1.4.1 Contexte

La mer Méditerranée Nord-Occidentale est une des seules régions dans le monde où la convection profonde hivernale peut avoir lieu. Deux éléments conditionnant la convection profonde y sont réunis : une faible stratification favorisée par l'intensification hivernale de la circulation cyclonique et des forçages atmosphériques intenses induisant un refroidissement et une évaporation importants des eaux en surface. La convection profonde participe à la circulation à l'échelle du bassin méditerranéen en renouvelant les masses d'eau de fond.

Au cours de la dernière décennie, l'augmentation des ressources en calcul a permis de nombreuses études sur la dynamique à sous-mesoéchelle et son impact sur les plus grandes échelles. En prenant la définition de McWilliams (1985), la sous-mesoéchelle fait référence à des échelles inférieures au rayon de déformation où l'influence de la rotation terrestre commence à diminuer, voire à devenir négligeable dans le cas de la turbulence 3D.

En mer Méditerranée Nord-Occidentale, le rayon de déformation est en moyenne inférieur à 10 km et atteint des valeurs plus faibles en hiver, d'environ 5 km jusqu'à des valeurs très faibles en cas d'absence de stratification. Pour représenter des échelles inférieures au rayon de

déformation dans un modèle de circulation, la résolution horizontale est fixée à 400m pour une résolution effective de l'ordre de sept fois la résolution, soit environ 3 km (Marchesiello et al, 2011).

Cette étude a pour but d'étudier le rôle des structures à sous-mésoéchelle dans la destruction de la stratification, le mélange intense et la phase de restratification après l'arrêt des forçages atmosphériques intenses. Plus précisément, les questions portent sur le rôle des fines échelles dans l'extension de la zone de convection, sur les échanges verticaux lors des épisodes de convection et sur la redistribution des masses d'eau après les épisodes de convection.

#### 6.1.4.2 Simulations numériques

Le modèle numérique est CROCO (<https://www.croco-ocean.org/>). Le contexte de modélisation reste dans le cadre de l'hypothèse hydrostatique. L'emprise et la bathymétrie de la configuration sont montrées sur la figure 6-4. Les dimensions sont de 3468 par 1363 mailles dans les directions zonale et méridienne respectivement, et la discrétisation verticale se fait sur 80 niveaux. Les simulations sont réalisées sur 18 nœuds en utilisant 486 cœurs pour CROCO et 18 cœurs pour la librairie XIOS qui gère l'écriture des résultats. Avec cette configuration, un mois de simulation est réalisé en une douzaine d'heures environ.

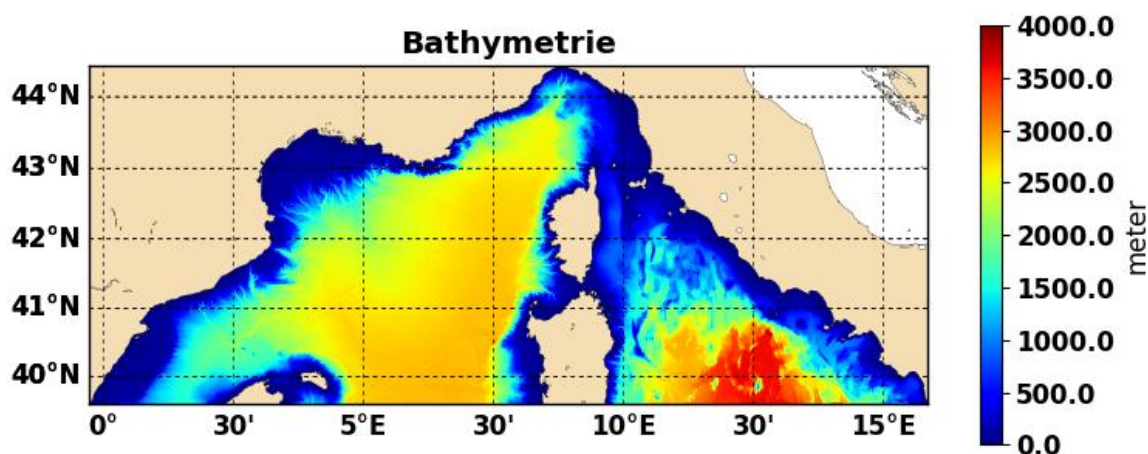


Figure 6-4 Bathymétrie de la configuration

Le modèle est intégré d'Août 2012 à Mai 2013 afin de reproduire l'intense événement de convection de l'hiver 2012/2013. La figure 6-5 montre un aperçu des vitesses verticales obtenues dans une zone où la convection a lieu. Les vitesses négatives atteignent des valeurs supérieures à 20 cm/s et sont concentrées dans des zones très fines séparées les unes des autres de quelques kilomètres. Ces ordres de grandeur sont concordants avec les valeurs calculées par Margirier et al (2017) à partir de données de gliders.

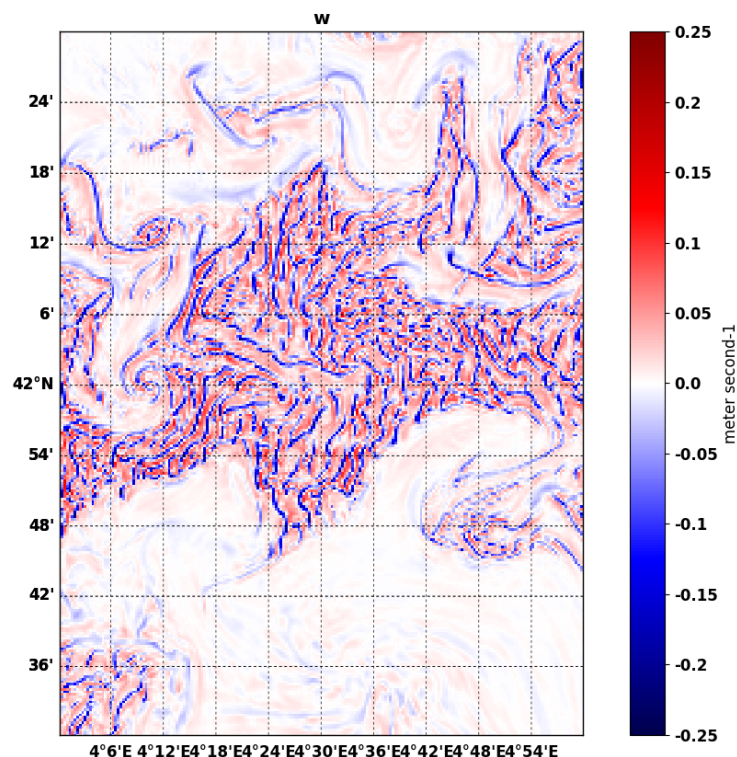


Figure 6-5 Vitesses verticales à 600 m de profondeur le 25 Février 2013

#### 6.1.4.3 Références

McWilliams, J. C. (1985). *Submesoscale, coherent vortices in the ocean*. *Reviews of Geophysics*, 23(2), 165-182.

Marchesiello, P., Capet, X., Menkes, C., & Kennan, S. C. (2011). *Submesoscale dynamics in tropical instability waves*. *Ocean Modelling*, 39(1-2), 31-46.

Margirier, F., Bosse, A., Testor, P., L'Hévéder, B., Mortier, L., & Smeed, D. (2017). *Characterization of convective plumes associated with oceanic deep convection in the northwestern Mediterranean from high-resolution in situ data collected by gliders*. *Journal of Geophysical Research: Oceans*, 122(12), 9814-9826.

#### 6.1.5 ETude des Ondes Internes au large des LandEs (ETOILE)

*Emeric Baquet (Ifremer-ODE-LOPS-OC)*

Le projet ETOILE consiste à étudier le devenir des ondes internes dans le coin Sud-Est du Golfe de Gascogne, au large des Landes, en particulier les paquets d'ondes internes solitaires, ou solitons, à haute fréquence (période inférieure à 20 minutes) et fine échelle (O(100m)). Ces solitons sont le résultat d'un équilibre entre la dispersion non hydrostatique et le raidissement dû aux non-linéarités, dans le cadre d'une colonne d'eau stratifiée. Une campagne de mesures a été réalisée entre le 9 et le 30 juillet 2017 et a mis en évidence la propagation de paquets de solitons sur le plateau des Landes. Ces paquets sont pour l'essentiel portés par la marée interne semi-diurne. C'est la première fois que ces solitons sont ainsi observés dans cette zone.

Dans ce contexte, une modélisation du Sud Golfe de Gascogne est réalisée, afin de reproduire ces phénomènes haute fréquence, en particulier localiser les zones de génération. La modélisation à de telles échelles spatio-temporelles nécessite une résolution fine et donc de grandes grilles d'où l'utilisation de Datarmor. Le modèle numérique utilisé est CROCO (pour Coastal and Regional Ocean COmmunity model, [www.croco-ocean.org](http://www.croco-ocean.org))

Dans un premier temps, une modélisation hydrostatique à 500m de résolution horizontale (866x745 points) et 40 niveaux verticaux est faite. Elle a pour but de représenter la marée interne porteuse. Plusieurs zones de génération, locale et lointaines sont identifiées (Figure 6-6), en raison de la bathymétrie particulière de la zone, en particulier le double plateau (Figure 6-7).

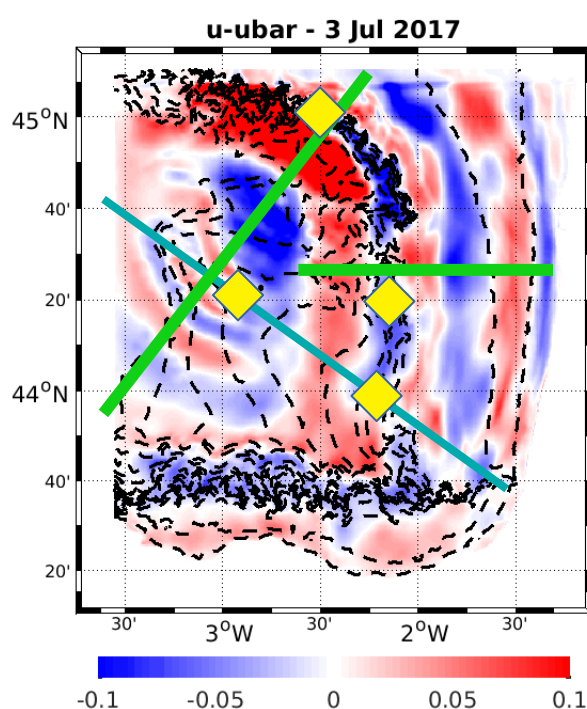


Figure 6-6 : Vitesse barocline (m/s) associée à la marée interne, à 50m de profondeur, en régime établi. Les losanges jaunes représentent les zones de génération, les radiales vertes et bleue les directions de propagation

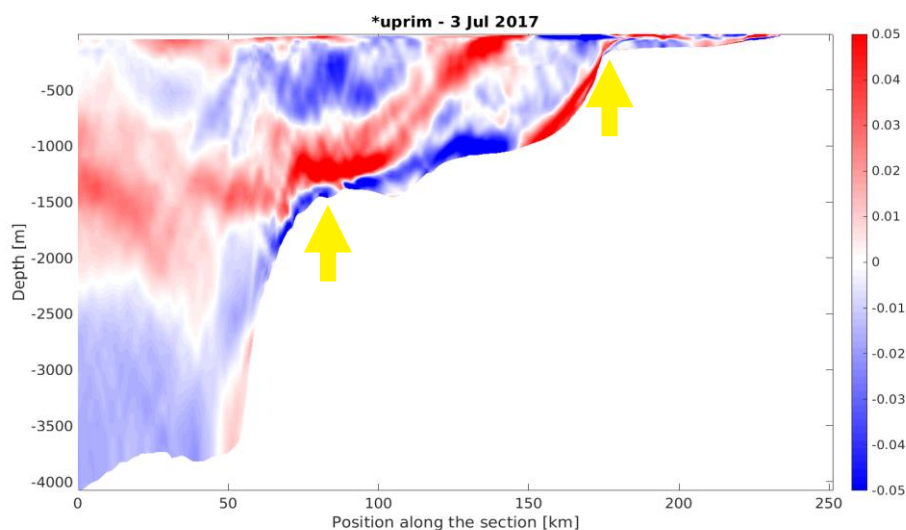


Figure 6-7 : Coupe verticale sur le plateau des Landes (radiale bleue Fig.1) de la vitesse barocline liée à la marée interne. On observe des rayons au niveau des zones de génération, symbolisées par les flèches jaunes.

Enfin, une modélisation non hydrostatique, à plus haute résolution horizontale, sur une plus petite emprise (100m, grille horizontale 873x1416points) avec 40 niveaux verticaux est en cours de réalisation. Les premiers résultats montrent la propagation de solitons sur le plateau (Figure 6-8).

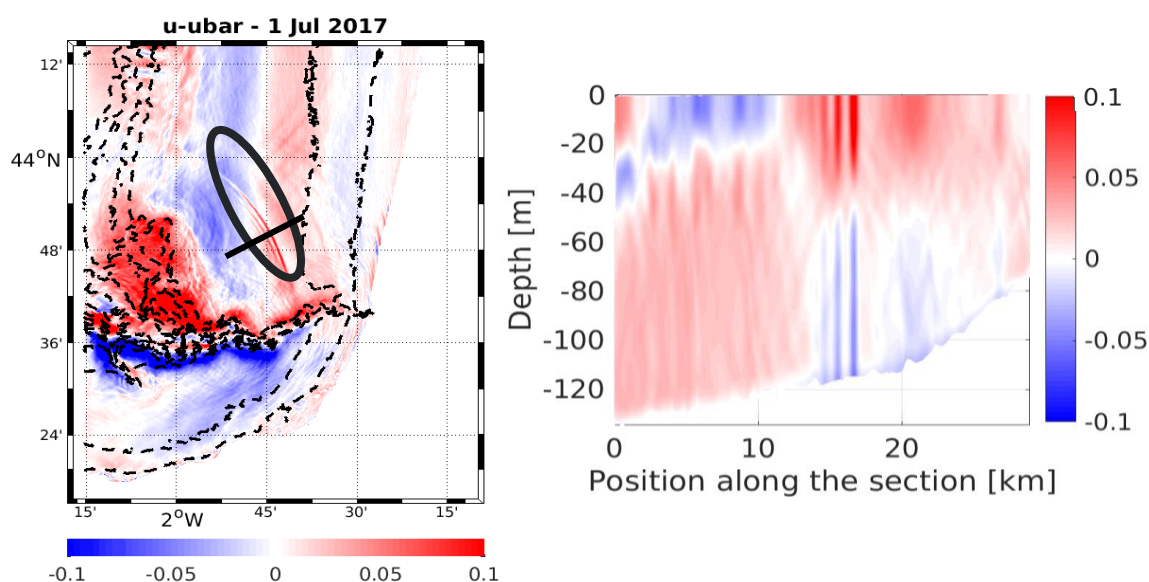


Figure 6-8 : Vitesse barocline (m/s) liée au passage des ondes internes à 20m de profondeur / coupe verticale de la vitesse barocline sur la radiale de propagation du paquet de solitons. On observe au passage des solitons un cisaillement vertical

## 6.1.6 Régionalisation des modèles de climat dans le Pacifique Sud

*Permanents* : Swen Jullien (Ifremer ODE-LOPS)

*Collaboration* : C. Dutheil (LOCEAN) / C. Menkes (LOCEAN) / Matthieu Lengaigne (LOCEAN) / Jérôme Vialard (LOCEAN) / Nicolas Jourdain (LGGE)

L'évolution des précipitations, et des événements extrêmes tels que les cyclones tropicaux, dans un contexte de réchauffement climatique global sont des enjeux sociétaux capitaux. Malgré de nombreux efforts, la modélisation climatique de ces phénomènes comporte encore aujourd'hui de fortes incertitudes, notamment dans le Pacifique Sud. Il a été montré que les modèles de climat globaux sont incapables de reproduire de manière réaliste les conditions océaniques et atmosphériques moyennes de cette région (e.g. double ITCZ et/ou langue d'eau froide équatoriale).

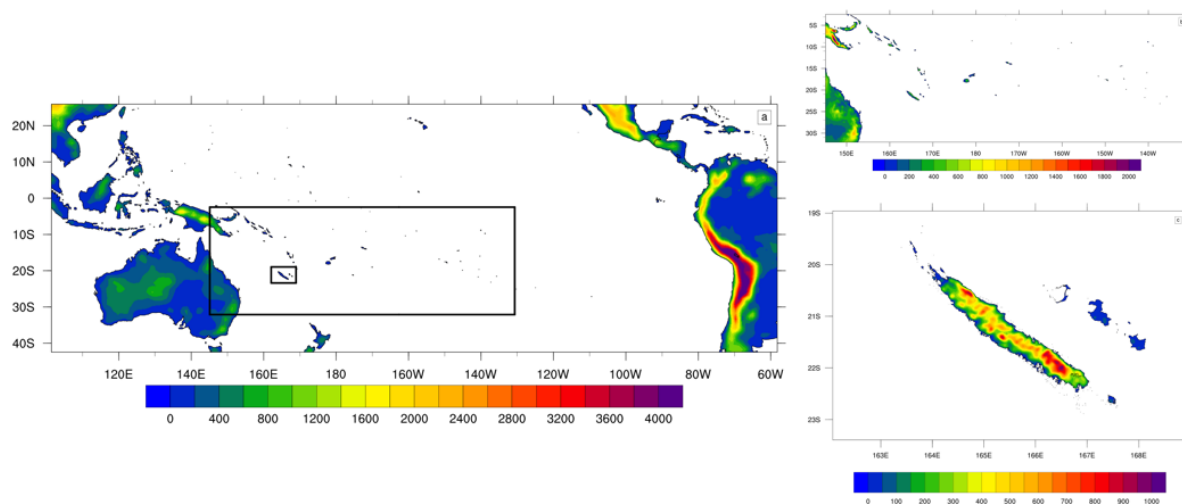


Figure 6-9 : Illustration des trois domaines de simulation

La descente d'échelle dynamique des modèles de climat globaux avec des modèles atmosphériques régionaux permet à la fois d'augmenter la résolution horizontale, et de réduire les biais sur la période historique. Ici nous avons modélisé trois domaines à des résolutions horizontales différentes (Pacifique à 1°, Pacifique sud à 1/5° et Nouvelle-Calédonie à 1/25°, Fiure 6-9) avec le modèle WRF.

Nous avons réalisé 12 simulations, résumées dans le tableau ci-après.

Tableau des simulations réalisées : chaque simulation a été réalisée dans 2 configurations du modèle atmosphérique représentant diverses paramétrisations physiques des effets sous-maille		
Nom Simulation	Période	Domaine (nom + résolution)
Présent (PD)	1980-2016	Pacifique [100°E-310°E;42°S-26°N] (1°) + Pacifique sud [145°E-230°E ;33°S-2°S] (1/5°)
Changement climatique (CC)	2072-2108	Pacifique [100°E-310°E;42°S-26°N] (1°) + Pacifique sud [145°E-230°E ;33°S-2°S] (1/5°)
Changement climatique avec correction de biais (COR)	2072-2108	Pacifique [100°E-310°E;42°S-26°N] (1°) + Pacifique sud [145°E-230°E ;33°S-2°S] (1/5°)
Présent (PD)	2080-2096	Nouvelle-Calédonie [162°E-172°E ;24°S-18°S] (1/25°)
Changement climatique (CC)	2080-2096	Nouvelle-Calédonie [162°E-172°E ;24°S-18°S] (1/25°)
Changement climatique avec correction de biais (COR)	2080-2096	Nouvelle-Calédonie [162°E-172°E ;24°S-18°S] (1/25°)

Bilan de la consommation : 47 To de stockage et 36.10<sup>6</sup>h CPU.

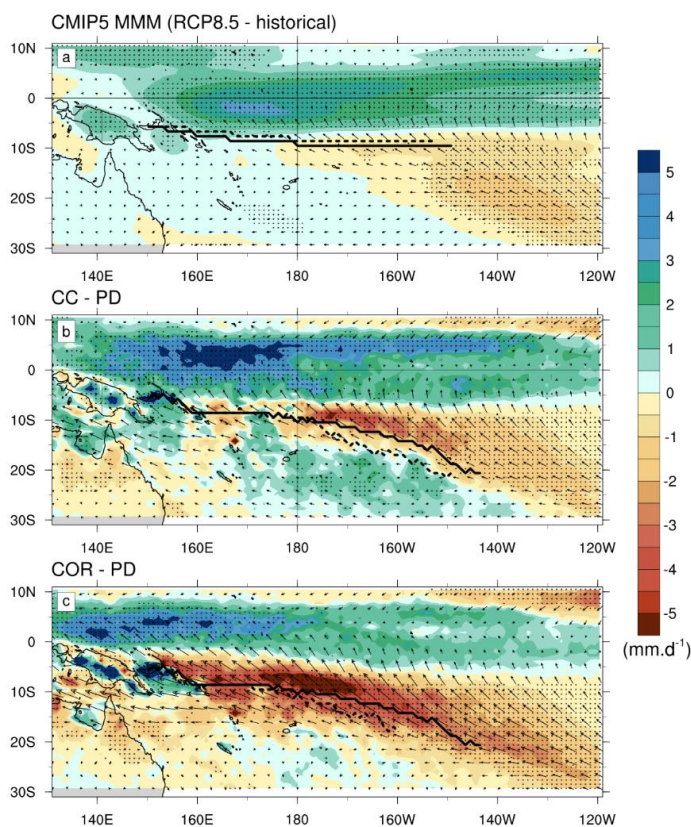


Figure 6-10 : Différences des précipitations (mm/jour, couleurs) et des vents de surface (m/s, vecteurs) en Décembre-Janvier-Février entre le futur et le présent ; (a) pour la moyenne multi-modèles des modèles de climat CMIP5, (b) pour les simulations WRF changement

### 6.1.6.1 Résultats :

Nos simulations climatiques montrent une forte diminution des précipitations (~25%; Figure 6-10c) dans le Pacifique Sud lorsque le pattern de réchauffement appliqué en surface est corrigé, tandis que les modèles CMIP5 et les simulations régionales « classiques » (sans correction) ne prévoient aucun changement significatif (Figure 6-10ab).

Nous avons également évalué la sensibilité de la cyclogenèse à ces biais, et nous montrons une diminution de -50 % de la cyclogenèse dans le Pacifique Sud lorsque le pattern de réchauffement appliqué en surface est corrigé (Figure 6-11b) alors que la diminution n'est que de -11 % sans correction (Figure 6-11a).

Ces 2 résultats montrent qu'il est crucial pour les études sur le climat futur du Pacifique Sud de prendre en compte les biais des modèles de climat et de les corriger.

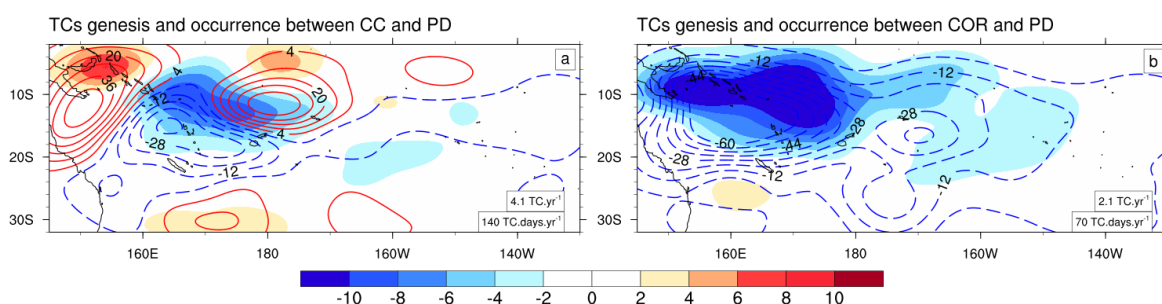


Figure 6-11 : Différence du nombre de cyclones (couleurs) et du nombre de jours cycloniques (contours) entre les simulations WRF futures et présentes (a) CC-PD, (b) COR-PD.

Enfin, les simulations à très haute résolution (cruciale pour bien représenter les reliefs, cf Figure 6-9) autour de la Nouvelle Calédonie ont révélé une diminution des précipitations de -20 % en moyenne, avec des contrastes spatiaux notamment entre la côte Est et la côte Ouest de la Nouvelle Calédonie. Le Sud et la côte Ouest de la Grande Terre sont les régions les plus impactées par ces changements avec une diminution des précipitations qui atteint -40 % à Nouméa.

L'ensemble de ces résultats ont été retrouvés dans les deux configurations physiques du modèle, confortant ainsi leur robustesse.

### 6.1.6.2 Publications :

Dutheil, C., Bador, M., Lengaigne, M., Vialard, J., Jullien, S., Jourdain, N., Lefevre, J., Pelletier, A., Menkes, C. Impact of surface temperature biases on climate change projections of the South Pacific Convergence Zone, *Clim. Dyn.* 2019

Dutheil, C., Bador, M., Lengaigne, M., Vialard, J., Jullien, S., Jourdain, N., Lefevre, J., Pelletier, A., Menkes, C. Impact of projected sea surface temperature biases on tropical cyclones projections in the South Pacific, *Soumis à Scientific Reports*



### 6.1.6.3 Conférences internationales :

Dutheil C., Bador M., Lengaigne M., Vialard J., Jullien S., Jourdain N., Lefevre J., Pelletier A., Sultan B., Menkes C. Impact of surface temperature biases on climate change projections of the South Pacific Convergence Zone and tropical cyclone activity. European Geosciences Union (EGU), Vienne, Avril 2019

Dutheil, C., Bador M., Lengaigne M., Vialard J., Jullien S., Jourdain N., Lefevre J., Pelletier A., Menkes, C. Impact of CMIP5 SST biases on the South Pacific Convergence Zone response to climate warming. International Conference for Southern Hemisphere Meteorology and Oceanography, Sydney, Australie, Fév. 2018

### 6.1.7 Mise en œuvre et performances du modèle CROCO pour les configurations cotières

*Matthieu Caillaud (Ifremer, ODE/DYNECO), Catherine Heyraud (Ifremer, ODE/DYNECO), Franck Dumas (SHOM)*

Le modèle CROCO (Coastal and Regional Ocean Community model, <http://www.croco-ocean.org/>) est pressenti comme étant la nouvelle plateforme de modélisation océanique partagée (IRD, l'INRIA, Ifremer, le Shom, l'Université de Toulouse et le CNRS) de référence dans les années futures. Une étude interne a porté sur la comparaison de ce modèle avec le modèle hydrodynamique MARS-3D utilisé historiquement par la communauté scientifique d'Ifremer pour différentes thématiques et notamment par le laboratoire DHYSED en hydro-sédimentaire (module MARS3D-MUSTANG).

Une mise en œuvre des deux modèles est réalisée de façon aussi identique que possible au cœur numérique près inhérent aux modèles (schéma temporel, schémas d'advection). Dans cette expérience jumelle, les deux modèles ont donc les mêmes grilles de calcul, les mêmes forçages (bathymétriques, atmosphériques, hydrodynamique initiale et aux frontières, fleuves), les mêmes paramétrisations physiques (frottement, mélange).

La zone d'étude est la baie de Seine, étudiée depuis de nombreuses années avec le modèle MARS ; Deux années (2015, 2016) ont été modélisées avec une résolution horizontale de 500m.

Les performances hydrodynamiques des deux modèles en Baie de Seine ont été évaluées grâce au large réseau de mesure déployé depuis quelques années dans cette zone (Figure 6-12 et Figure 6-13).

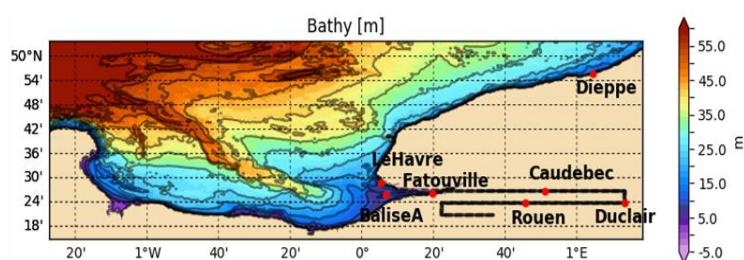


Figure 6-124 Localisation des marégraphes

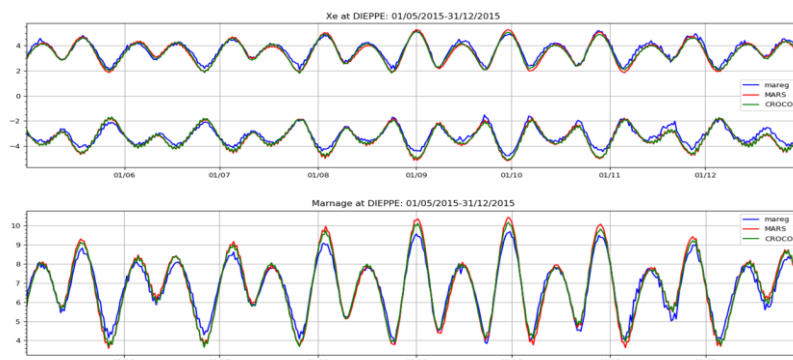


Figure 6-13 Evolution temporelle des enveloppes minimale et maximale des hauteurs d'eau (haut) et du marnage pour les marégraphes (bleu), MARS (rouge), CROCO (vert) pour 2015 à Dieppe

Pour cette inter-comparaison, le calculateur Datarmor a été utilisé pour produire un rejeu de deux ans pour les deux modèles. Afin de comparer les performances des modèles en termes de temps de calcul, les simulations ont été réalisées en MPI sur 11 cœurs (quatre nœuds de la grappe de calcul HPC).

Pour les traitements post-simulation (validation, comparaison des modèles et exploitation des résultats), de nombreux scripts python ont été soumis sur Datarmor cette fois sur les nœuds avec beaucoup de mémoire vive (cluster SMP). Enfin pour l'hébergement des résultats de modélisation les espaces SCRATCH et DATAWORK ont été indispensables à la réalisation de l'étude.

### 6.1.8 Projet DIETE

#### *Martin Plus – DYNECO Pélagos*

Diagnostic étendu de l'eutrophisation en baie de Vilaine - Rôle des sédiments dans le cycle des nutriments. Les actions de modélisation réalisées avec l'aide du calculateur DATARMOR concernent le développement d'un nouveau module ECOMARS de calcul des processus diagénétiques dans le sédiment. Les processus biologiques sédimentaires sont intégrés au module de dynamique sédimentaire MUSTANG déjà couplé à MARS. L'objectif est d'améliorer la simulation des cycles biogéochimiques dans la zone côtière afin de progresser sur la compréhension du rôle des sédiments dans le phénomène d'eutrophisation.

Le module, déjà codé, est actuellement testé sur une configuration 1DV (simulations séquentielles) forcée par le modèle MARS Loire-Vilaine à 800 m (grille de calcul 182x132 mailles, simulations annuelles, walltime 24 h, parallélisation MPI 28 cœurs, mémoire

demandée 6Gb). L'étape suivante sera le passage du 1DV vers le 3D, toujours sur la configuration Loire-Vilaine avec couplage MARS-MUSTANG-ECOMARS.

### 6.1.9 Projet Alex-Breizh

*Thèse Samuelson Nzeneri*

Les proliférations récurrentes des microalgues toxiques en zone côtière posent un problème sanitaire et sociétal important. L'espèce toxique *Alexandrium minutum* prolifère sur le littoral breton depuis plusieurs années, notamment dans les estuaires riches en nutriments (Penzé, Morlaix, Rance, Abers et plus récemment la rade de Brest). Cette espèce est l'enjeu de questions de recherche générale:

- Pourquoi une espèce devient-elle dominante à un moment et à un endroit donné dans un environnement où les facteurs de contrôles sont nombreux?
- Comment analyser les niches d'*A. minutum* et celles des espèces phytoplanctoniques associées ?
- Quelles sont les caractéristiques des habitats (hydrodynamisme, richesse en nutriments...) et quels traits physiologiques peuvent expliquer la répartition spatiale ?
- Y a-t-il un lien entre les différents blooms de la rade (connectivité) ou sont-ils indépendants géographiquement ?
- Qu'est-ce qui contrôle la dynamique d'*A. Minutum*?

L'approche retenue est la modélisation numérique couplant un modèle hydrodynamique de la rade et un modèle biologique qui intègre la diversité biologique et la compétition interspécifique pour les ressources. Le modèle numérique prend en compte, de manière la plus objective possible, la compétition pour les ressources nutritives entre *A. minutum* et une soixantaine d'autres espèces phytoplanctoniques comme facteur potentiel de régulation. Cette technique de modélisation est basée sur une approche en traits physiologiques pour représenter la communauté phytoplanctonique dans sa globalité.

Le modèle est codé en Fortran, en 2018 de nombreuses simulations ont été lancées sur DATARMOR, dans une version adimensionnelle (0D), de façon à tester différentes paramétrisations sur une période assez longue de 2009 à 2017. Les simulations sont effectuées en séquentiel. L'étape suivante consistera à spatialiser le modèle (3D) pour simuler la Rade de Brest dans une configuration MARS de résolution 250m (grille 191x109 mailles, MPI2D).

### 6.1.10 Rôle de l'adaptation locale et de la connectivité dans la structure et le fonctionnement des populations de coquilles Saint Jacques (*Pecten maximus*) en Rade de Brest, Mer d'Iroise et Manche

*William HANDAL (Thésard UMR LEMAR/Panorama/Génétique des Populations)*  
*Responsables : Pierre Boudry (Ifremer) et Grégory Charrier (UBO)*

### 6.1.10.1 Contexte scientifique du projet

La coquille Saint Jacques (*Pecten maximus*) est un bivalve à cycle benthopélagique fortement exploité le long des côtes Européennes. Au sein des eaux territoriales Françaises, sont retrouvés des gisements importants tels que la Baie de Seine, Saint-Brieuc, Granville et dans une moindre mesure la Baie de Morlaix et la Rade de Brest. La motorisation des engins de pêches au milieu du XX<sup>ème</sup> siècle associée ; à l'hiver rigoureux de 1962, aux blooms de *Gymnodinium cf nagasakiense* et à la compétition interspécifique avec

*Crepidula fornicata* a entraîné le déclin du gisement de Rade de Brest. En 1950, les débarquements étaient de 1500-2600t pour 100-300t en 1964. En 1983, un programme d'ensemencement dirigé par l'écloserie du Tinduff à vu le jour et entraîné une légère augmentation de la production de coquilles en Rade. Malgré le déclin du stock, et l'ensemencement, les études de génétique des populations basées sur les microsatellites n'ont révélé aucune différence de diversité génétique entre la Rade de Brest et les autres gisements exploités de Manche. L'hypothèse de maintien de la diversité génétique par des flux de gènes entre la rade de Brest et l'extérieur ont été avancées. Cependant des études de modélisation de dispersion larvaire ont décrit la Rade de Brest comme semi-fermée, la dispersion apparaissant limité entre la Rade de Brest et la Manche Ouest.

### 6.1.10.2 Problématique de travail

Est ce que la Rade de Brest apparaît isolée de la Mer d'Iroise et de la Manche Est ? Quelle est la structure génétique neutre et adaptative entre les gisements considérés ?

### 6.1.10.3 Utilisation de Datarmor

Afin d'étudier la diversité génétique au sein des gisements nous avons analysé le génôme via la technique de Genotype By Sequencing, afin d'identifier des SNPs, marqueurs de diversité génétique. Des centaines de millions de séquences ont été obtenues et décryptées via Datarmor à l'aide des logiciels suivant :

- Trimmomatic (Nettoyage des séquences)
- Cutadapt (Nettoyage des séquences)
- Stacks (Création d'un génôme de novo)
- VCFtools (Nettoyage des données SNPs)
- Plink (Identification des déséquilibre de liaison et de l'apparentement entre individus)
- Flk (Identification des SNPs sous sélection)

### 6.1.10.4 Devenir du projet

Grâce à l'utilisation de cette précieuse ressource bio-informatique qu'est Datarmor et les compétences des services bio-informatique d'Ifremer nous avons été en mesure de mener à bien notre projet donnant lieu à une prochaine publication.

### 6.1.11 Projet FOREVER : Modélisation de la dispersion larvaire en rade de Brest

*Sébastien Petton, Stéphane Pouvreau, Aline Brossin (Ifremer LEMAR PFOM-LPI)*

Cette étude a été réalisée dans le cadre du projet FEAMP "FOREVER". Ce projet français s'intègre dans un dispositif européen de plus grande ampleur dont l'objectif est de favoriser le retour de l'huître plate, espèce native des côtes européennes en voie de disparition. Le transport et la dispersion larvaire au sein de la rade de Brest ont été simulés à l'aide du modèle hydrodynamique MARS 3D (*Model for Applications at Regional Scale*) qui résout les équations de Navier-Stokes sous hypothèse hydrostatique (Lazure et Dumas, 2008) et permet une prise en compte fine des forçages et des caractéristiques hydrologiques des zones côtières.

Le modèle 3D de la rade de Brest s'étend de 48.20° à 48.44° N de latitude et de 4.09° à 4.72° W de longitude. Sa résolution horizontale est de 50 m et sa résolution verticale  $\sigma$  est de 20 niveaux. Cela représente une grille de 946 par 543 mailles. Il a été validé sur différents jeux de données datant de 2010 à maintenant (Petton et al., non publié). La marée est transmise grâce à un emboîtement de grille gigognes avec les constantes harmoniques en élévation du SHOM (produit CST France, 2004) qui sont utilisées aux limites d'un plus large modèle 2D qui recouvre la mer d'Iroise et le Finistère. Les paramètres atmosphériques viennent du modèle AROME de Météo-France (2.5 km de résolution horizontale). Les débits des trois principales rivières de la rade obtenus *via* la banque Hydro (<http://www.hydro.eaufrance.fr/>) sont pris en compte dans le modèle et corrigés du rapport de bassin versant (l'Aulne, l'Elorn, La Mignonne). Enfin, les conditions océaniques 3D de température et de salinité sont issues à une fréquence horaire d'une autre maquette MARS 3D qui a une résolution de 500 m centré sur le Finistère. Il en est de même lors de l'initialisation du modèle pour le *spin-up* qui dure 30 jours (période estivale).

Pour simuler le transport larvaire, une approche eulérienne a principalement été utilisée pour étudier l'influence des marées, du vent et du comportement larvaire. Ainsi différents comportements de nage assez distincts ont été testés (5 cas représentés dans la **figure 6-14**) : larves ayant un comportement neustonique (nage en surface, **figures 6-14.a et 6-14.b**), suprabenthique (nage proche du substrat, **figures 6-14.d et 6-14.e**) ou intermédiaire (**figure 1.c**). Enfin, un sixième cas (de référence) a aussi été testé : larves sans comportement, considérées comme des particules inertes avec une flottabilité neutre. Ces comportements résultent de la combinaison linéaire de paramètres physiologiques tel que la vitesse de nage, la vitesse de chute et la période d'activité qui évoluent avec la vie larvaire.

Deux périodes différentes : une au début d'une période de vives eaux (du 19/07/2016 au 29/07/2016) avec des forts coefficients durant les premiers jours et une autre au début d'une période de mortes eaux (du 25/07/2016 au 04/08/2016)). Durant ces périodes, le vent est un vent faible d'Ouest. L'impact du vent étant faible sur la dispersion larvaire en rade de Brest (Alimou, 2017), cet objectif a été abordé en second lieu. Par contre, 13 moments d'émission des larves au cours d'un cycle de marée espacée d'une heure ont été simulés. Pour

l'interprétation des cartes de dispersion larvaire, les concentrations larvaires obtenues sont moyennées sur un cycle de marée.

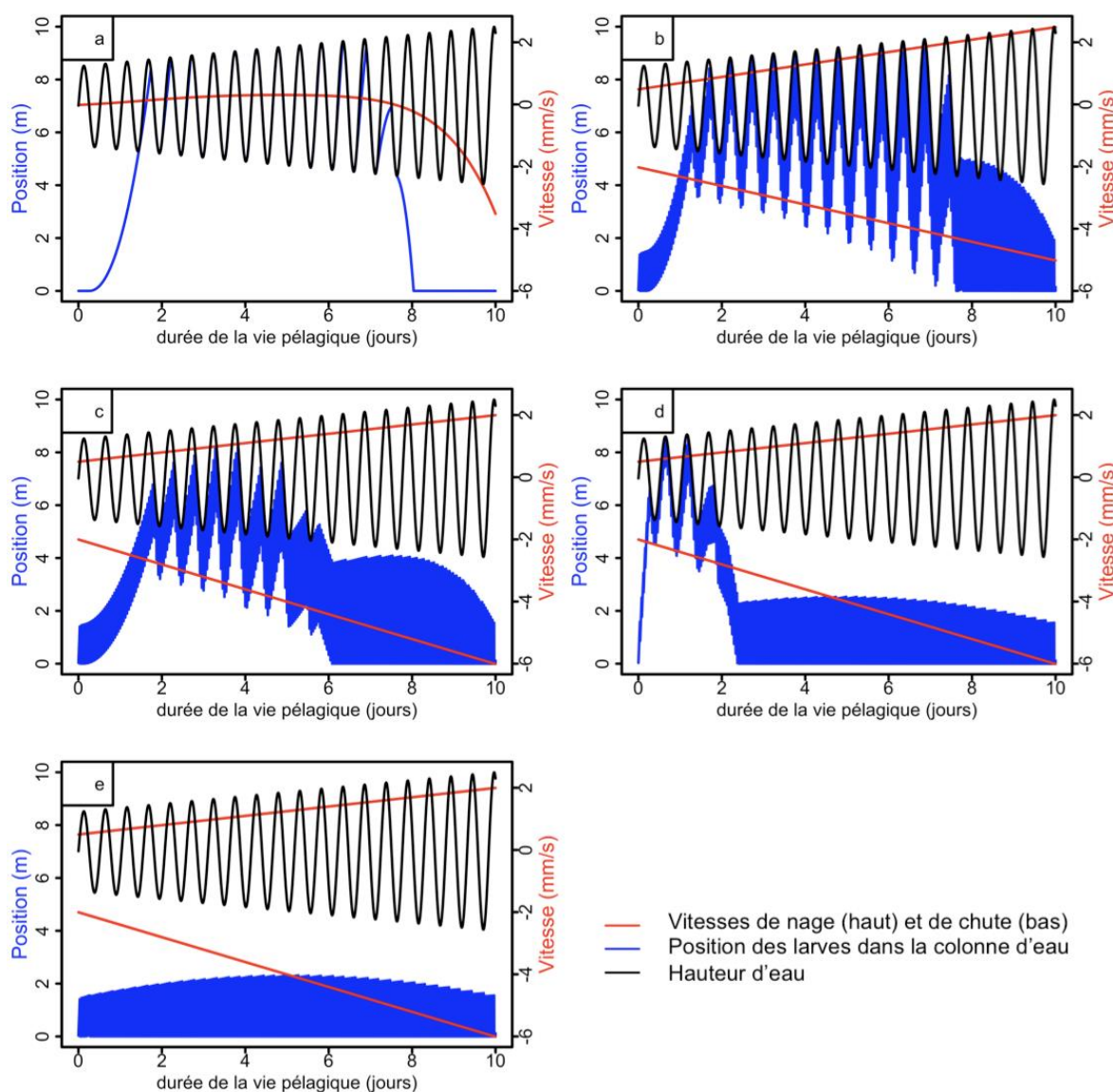


Figure 6-14 : Différents comportements potentiels de larves ici dans une colonne d'eau 1DV subissant la marée. a) comportement neustonique fort ; b) comportement neustonique modéré ; c) comportement intermédiaire (neustonique puis suprabenthique); d) comportement suprabenthique modéré ; e) comportement suprabenthique fort

Sur la **figure 6-15**, il ressort clairement que les comportements neustoniques sont plus dispersifs que les comportements suprabenthiques, que les émissions soient en mortes eaux (non montrés ici) ou en vives eaux. Quel que soit le comportement et principalement en mortes eaux, les abondances larvaires après 10 jours de dispersion sont plus importantes dans certaines zones de la rade. Les larves ayant un comportement suprabenthique fort ou modéré sont très peu dispersées en dehors de la rade et de la rade Sud, plus de 75% des larves émises étant conservées au sein de la rade Sud. Au contraire, les larves n'ayant pas de comportement ou un comportement autre que suprabenthique sont beaucoup plus dispersées en dehors de la rade et de la rade Sud : moins de 50% des larves sont conservées au sein de la rade Sud.

Cette étude de modélisation a été suivie d'une campagne de validation *in-situ* du recrutement sur une période de 15 jours. Cette validation a permis la sélection d'un comportement parmi les différents proposés. D'autres campagnes de validation sont prévues.

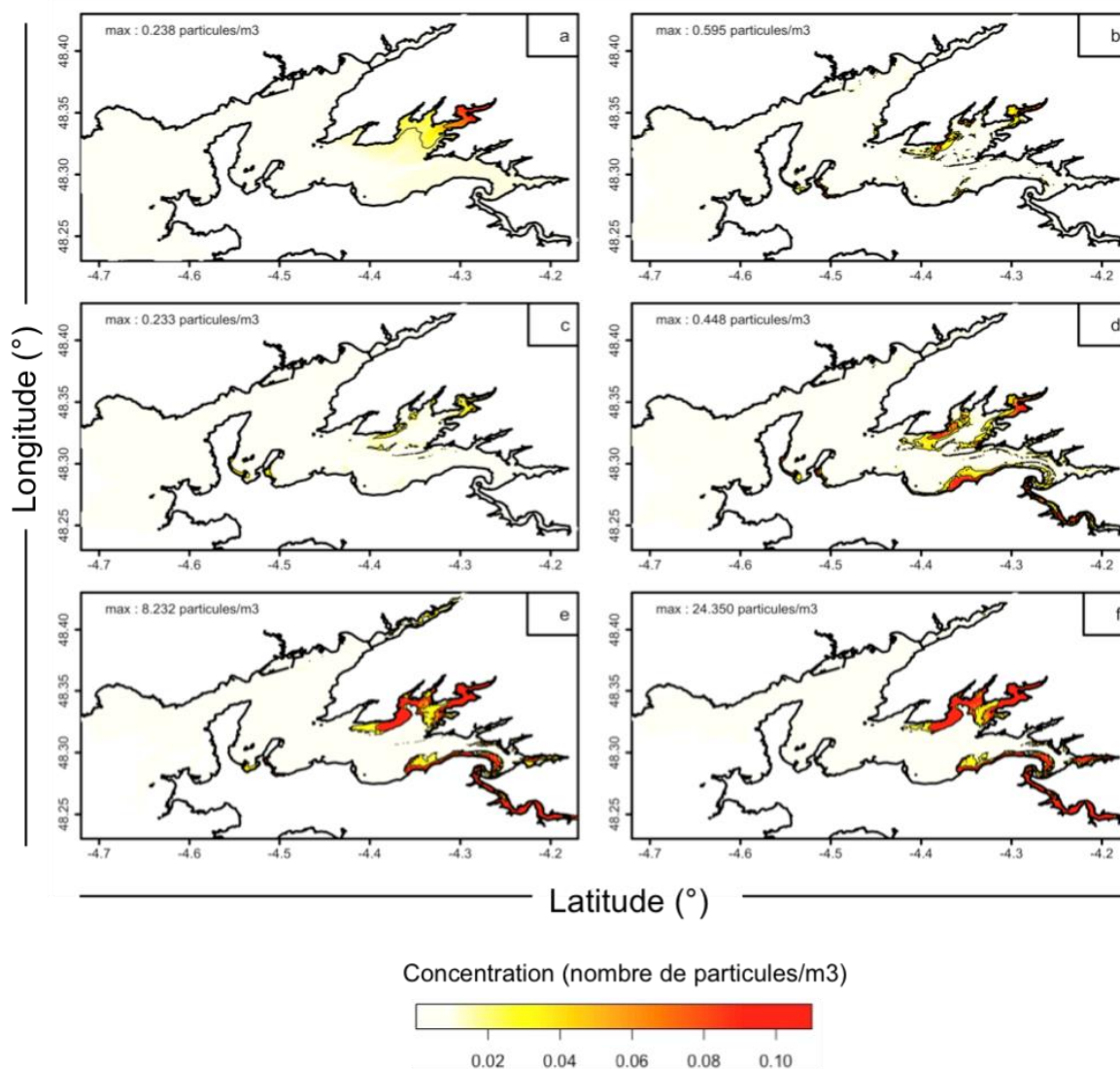


Figure 6-15 : Dispersion des larves émises ( $n=30$  millions) au banc du Roz après 10 jours de vie pélagique en période de vives eaux, en fonction des différents comportements : a) sans comportement ; b) comportement neustonique fort ; c) comportement neustonique modéré ; d) comportement intermédiaire ; e) comportement suprabenthique modéré ; f) comportement suprabenthique fort. Les concentrations représentées sont les moyennes des concentrations sur le dernier cycle de marée

Aspects numériques sur Datarmor :

- 336 processeurs avec une compilation hybride OMP + MPI
- 1 run tourne en 20h
- 13 runs par scenario avec un peu plus de 12 scenarii à l'état initial de l'étude.
- 7 substances émises par run

### 6.1.12 OSMOSE-MED : un modèle écosystémique intégré pour comprendre et gérer les ressources d'une Méditerranée sous pressions

*Fabien Moullec, Laure Velez, Nicolas Barrier, François Guilhaumon et Yunne-Jai Shin (UMR MARBEC), Philippe Verley (UMR AMAP) et Caroline Ulses (Laboratoire d'Aérodynamique, Toulouse).*

La Méditerranée représente l'un des plus grands réservoirs de biodiversité à l'échelle mondiale. Plus qu'un point chaud de biodiversité, celle-ci se transforme progressivement en un point chaud du changement global sous les effets combinés de la surexploitation des ressources marines et du changement climatique. Pour être en mesure d'émettre des avis scientifiques sur la stratégie à adopter pour une gestion écosystémique des ressources à l'échelle du bassin méditerranéen, il existe un véritable défi scientifique à développer nos capacités de représentation explicite des dynamiques spatiales multi spécifiques sous l'influence combinée de la variabilité climatique et de la pression de pêche.

En dépit d'un grand nombre de modèles développés localement ou régionalement, aucun travail de modélisation n'a envisagé de projeter des modifications des assemblages d'espèces à l'échelle du bassin méditerranéen avec une modélisation explicite intégrée (à la fois multi spécifique, spatiale, trophique, basée sur les traits de vie, représentant le cycle de vie entier des espèces) des dynamiques de la richesse spécifique. Dans cet objectif, un modèle end-to-end, s'appuyant sur le couplage d'un modèle de hauts niveaux trophiques OSMOSE<sup>8</sup> à des modèles de physique et de biogéochimie NEMOMED-12 et ECO3M-S, a été développé : OSMOSE-MED. Il s'agit du premier modèle trophique de ce type intégrant une aussi grande richesse spécifique et sur une échelle spatiale aussi vaste.

Ce modèle a été calibré puis utilisé sur la machine de calcul Datarmor. Il a donné lieu à trois publications, une en révision, une soumise et une en préparation.

#### *Bibliographie*

*Catching the big picture of the Mediterranean Sea biodiversity with an end-to-end model of climate and fishing impacts.* En révision pour Progress in Oceanography.

*An End-to-End model revealed losers and winners in a warming Mediterranean Sea.* Soumis à Frontiers in Marine Science.

*Rebuilding Mediterranean marine resources under climate change.* En préparation.

### 6.1.13 Le modèle APECOSM

Le modèle APECOSM est un modèle écosystémique spatialisé et structuré en taille, qui simule de manière mécaniste les différents flux d'énergie intervenant au sein des écosystèmes marins. Plusieurs travaux de développement ont été entrepris sur la machine de calcul Datarmor.

---

<sup>8</sup> OSMOSE est une plate-forme de modélisation écosystémique conçue sous la forme d'un modèle individu-centré pour simuler la dynamique des communautés de poissons. L'hypothèse centrale d'OSMOSE est que la prédation est un processus opportuniste qui dépend d'une part de la taille relative proie / prédateur et d'autre part de la probabilité qu'un prédateur croise sa proie à un endroit donné et un instant donné.



### 6.1.13.1 Couplage du modèle Apecosm avec le modèle NEMO-PISCES *Nicolas Barrier (IRD, MARBEC), Olivier Maury (IRD, MARBEC) et Olivier Aumont (IRD, LOCEAN)*

En prévision de plusieurs projets à venir (projet ANR CIGOEF par exemple), un couplage du modèle écosystémique Apecosm (codé en C/C++) avec le modèle physique biogéochimique NEMO-PISCES (codé en Fortran) est en cours de développement. L'objectif de ce couplage est d'estimer les rétroactions des niveaux trophiques supérieurs sur la biogéochimie (via la prédation, l'excrétion et l'égestion), ainsi que d'estimer la contribution du mélange induit par les mouvements des poissons (< biomixing >) sur le mélange océanique total. Et ceci dans le cadre de simulations historiques ou de scénarios de changement climatique (scénarios RCP) effectués dans le cadre de CMIP6.

L'intérêt de ce couplage est de bien cloisonner les développements d'Apecosm (développé au sein de l'UMR MARBEC) et de NEMO-PISCES (développé au LOCEAN), et ainsi de bénéficier, à moindre coût, des dernières versions de chaque modèle dans ces simulations couplées.

Ce couplage est désormais fonctionnel, à la fois en one-way (NEMO-PISCES => Apecosm) et en two-ways (NEMO-PISCES ⇔ Apecosm). Cependant, les résultats de ce couplage obtenus à partir d'une simulation globale sur la grille ORCA2, ne sont pas encore satisfaisant. En effet, la prédation des espèces Apecosm envers certaines espèces simulées par PISCES est trop forte, ce qui induit un changement d'équilibre de la biogéochimie non conforme aux observations. Pour améliorer ces résultats, de nombreux tests sont effectués sur les paramètres des modèles NEMO-PISCES et Apecosm, ce qui nécessite beaucoup de ressources de calcul et de stockage.

### 6.1.13.2 Couplage du code Apecosm avec un modèle de pêche.

Dans le cadre du projet ANR CIGOEF, un modèle de pêche a été développé et couplé avec le modèle écosystémique Apecosm. Ce couplage a d'abord été testé en local sur une configuration idéalisée (configuration GYRE). Il est désormais en cours de test sur une configuration globale de type ORCA2, nécessitant le recours à la machine de calcul Datarmor. Une fois ce couplage valide, l'objectif sera de faire tourner plusieurs scénarios couplés écosystèmes/pêches en utilisant des sorties hydrodynamiques provenant de scénarios de changement climatiques (sorties IPSL-CM5).

## 6.1.14 Divers travaux de l'unité MARBEC

### 6.1.14.1 Inventaires de biodiversité marine par des approches basées sur l'ADN environnemental

*Miriam Brandt (Ifremer, Marbec), Blandine Trouche (Ifremer, Marbec), Laure Quintric (Ifremer, IRSI/BIOINFO), Sophie Arnaud-Haond (Ifremer, Marbec)*

Dans le cadre du projet Pourquoi Pas les Abysses piloté à Marbec, Datarmor a été utilisé pour développer et utiliser un pipeline dédié aux approches de métabarcodage environnemental sur 5 fragments de gènes, avec pour objectif de cibler l'ensemble des procaryotes et des

eucaryotes présents notamment dans les communautés benthiques du plancher océanique. Ce pipeline développé dans le cadre de la thèse de Miriam Brandt, avec la participation de Blandine Trouche (également doctorante sur le projet pourquoi pas les Abysses), en collaboration avec la cellule bio-informatique de l'Ifremer, fonctionne maintenant est utilisé par l'ensemble des équipes de Sète, Brest et Roscoff impliquées dans le projet.

#### 6.1.14.2 Evaluation des inférences par EM de modèles hiérarchiques HMM et HSMM

*Nicolas Bez (IRD, Marbec) et Stéphanie Mahévas (Ifremer Nantes, EMH)*

La machine de calcul Datarmor est utilisée pour tester la qualité des inférences produites par les algorithmes Espérance-Maximisation (E-M) dans les cas des modèles hiérarchiques utilisés en analyses de données de tracking. Ce sont des algorithmes itératifs nécessitant deux phases de calcul à chaque itération. La qualité des inférences est évaluée dans un processus de Simulation-Estimation portant sur 12 cas d'étude différents (3 bateaux et 4 niveaux de fréquence d'acquisition des données). Les inférences sont également analysées quand la structure du modèle utilisé pour l'estimation des paramètres est différente de la structure du modèle utilisé pour simuler les données.

Toutes ces expériences sont parallélisées et tirent parti à plein des moyens de calcul de Datarmor.

#### 6.1.14.3 Comparaison de trajectoires de drifters observées et simulées par le modèle Ichthyop

*Hilaire Amemou (IRD, Marbec), Christophe Lett (IRD, Marbec), Nicolas Barrier (IRD, Marbec)*

Afin de mieux comprendre la dispersion et la croissance des larves de sardinelles dans le Golfe de Guinée, des simulations lagrangiennes sont effectuées avec le modèle de dispersion larvaire Ichthyop.

Dans le but d'appréhender la pertinence des trajectoires simulées, ces dernières sont comparées à des trajectoires observées de drifters. Pour chaque trajectoire observée, une simulation Ichthyop est effectuée sur la période correspondant à la durée de vie du drifter. Cet exercice est ensuite répété en utilisant différents forçages hydrodynamiques (OSCAR, GlobCurrent), de résolution spatiale (1=3, 1=8) et temporelle (journalière, 3h) différentes.

Le grand nombre de simulations Ichthyop ainsi que le grand volume de données limite l'utilisation des machines locales. Ces simulations sont donc désormais effectuées sur Datarmor, avec comme objectif à terme de faire tourner ces simulations en parallèle, et non plus en séquentiel.

### 6.1.15 Projet PRESENCE : PRESsions sur les Ecosystèmes récifo-lagonaires de Nouvelle-Calédonie

*Romain Le Gendre (IFREMER LEAD-NC, Nouméa, Nouvelle-Calédonie)*

*Collaboration : T rence Desclaux (IFREMER Noum a), No mie Lalau (IFREMER Noum a), S bastien Petton (IFREMER PFOM-LPI Argenton), Benoit Soulard (IFREMER Noum a)*

Depuis de nombreuses ann es, la Nouvelle-Cal donie est reconnue comme un des points chauds de la biodiversit  mondiale, h bergeant une des richesses end miques les plus  lev es au monde (Kier et al., 2009). Elle abrite par ailleurs le plus grand complexe corallien au monde (plus de 7000 km<sup>2</sup>) et pas loin de 75% de la surface des r cifs et lagons des territoires fran ais. L' tat exceptionnel de sa biodiversit  marine lui a valu en 2008 l'inscription d'une partie de ses lagons au Patrimoine Mondial de l'Humanit . Pourtant, l'industrie mini re, tr s pr sente, a fa onn  depuis plus d'un si cle les paysages terrestres et littoraux. L'urbanisation croissante et l'exploitation agricole induisent  galement des pressions de plus en plus soutenues sur les  cosyst mes coralliens, auxquelles viennent se surajouter les impacts du changement global. Dans ce contexte, un des objectifs du projet PRESENCE s'attachait   la mise en place d'une strat gie de mod lisation repr sentant d'une mani re r aliste le continuum terre-lagon-oc an, afin de mieux appr hender la dynamique et le devenir des apports terrig nes au sein des lagons cal doniens.

Concernant les apports terrestres, le mod le GR4H, un mod le pluie-d bit au pas de temps horaire a  t  d ploy  sur Datarmor pour repr senter les bassins versants cal doniens (Desclaux et al., 2018), caract ris s principalement par des crues tr s intenses et rapides. Apr s la mise en place d'une proc dure de calibration idoine, ce mod le s' st av r  efficace pour simuler les d bits des rivi res cal doniennes et fournit d sormais les flux liquides au mod le MARS3D d velopp  sur les lagons. Datarmor a notamment  t  utilis  pour permettre la proc dure de param trisation du mod le GR4H sur de nombreux bassins et une grande p riode temporelle (Algorithme de Nelder-Mead et k-fold partitioning).

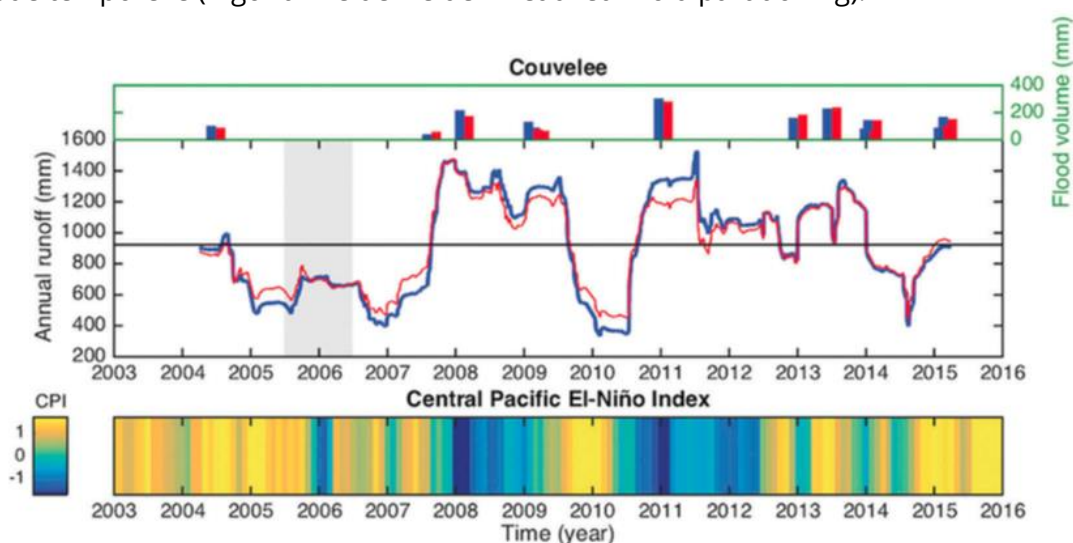


Figure 6-16 :  coulements annuels pour la rivi re de la Couvel e issus des observations (bleu) et des simulations versus CPI (Central Pacific El-Ni o Index). La moyenne sur la p riode d' tude est repr sent e par la ligne noire. Les volumes des crues principales sont visibles sur les histogrammes situ s au-dessus du graphique principal.

Une maquette hydrodynamique MARS3D très conséquente a été développée avec l'objectif de représenter la circulation régionale et également d'intégrer les lagons de la Grande Terre à haute résolution. Pour cela une configuration AGRIF, gourmande en temps de calcul (1500m puis 300m, 50 niveaux sigma généralisés) a été mise en place. En termes de forçages, celle-ci est forcée par FES2012 pour la marée, PSY2V4R4 (MERCATOR) pour les conditions océaniques, ALADIN ou AROME pour l'atmosphère, et GR4H donc pour les apports terrestres. Suite à une classification des séquences débits-vents post-crués, les configurations typiques de panaches autour du Grand Nouméa ainsi que leur dynamique ont par exemple été analysés et confrontés aux observations in-situ. De nouveaux développements sont en cours pour intégrer les flux entrants dans les lagons par déferlement sur les récifs barrières (forçage WW3) et le projet PRESENCE va s'attacher par la suite à fournir des indicateurs du fonctionnement hydrodynamique à l'échelle du territoire calédonien, notamment lors d'épisodes météorologiques intenses (e.g cyclones, dépressions tropicales).

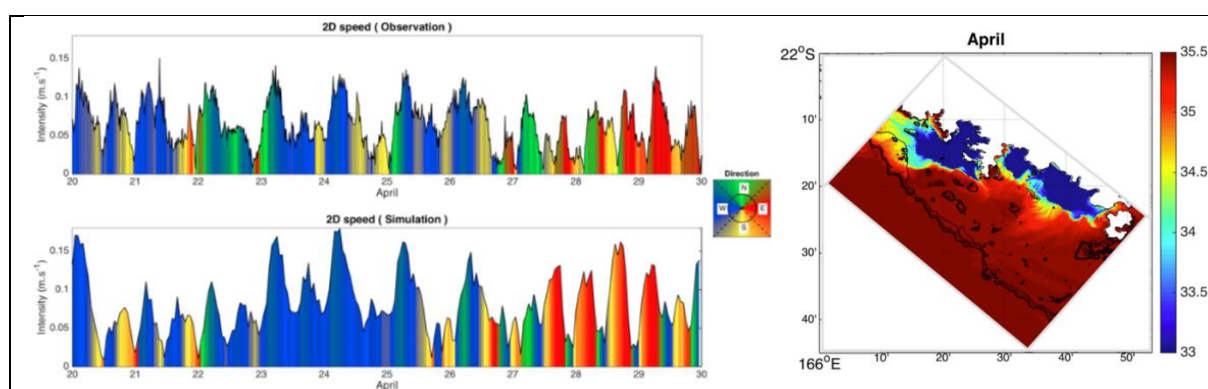


Figure 6-17 : Gauche : Comparaison du courant observé (haut) à la station Moïse et du courant simulé par la maquette MARS3D au même point. Droite : Salinité minimale simulée suite au passage du cyclone Cook (Avril 2017).

## 6.1.16 Étude de sensibilité de modèles hydrologiques pour la prévision des crues dans un estuaire

*Anne-Laure Tiberi-Wadier (Cerema), Nicole Goutal (EDF R&D), Sophie Ricci (CERFACS), Philippe Sergent (Cerema)*

### 6.1.16.1 Contexte et objectifs de l'étude

Afin de réaliser les cartes de vigilance bi-quotidiennes, le SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) et les 19 SPC (Services de Prévision des Crués) utilisent entre autres des résultats de modèles numériques généralement lancés de manière déterministe (prévisions de pluie et de débit hydrologique ou hydraulique), c'est-à-dire avec une combinaison unique de données d'entrée et de modèles. La chaîne de prévision étant sujette à de nombreuses incertitudes (données météorologiques observées et prévues, données hydrologiques, modèles hydrologiques ou hydrauliques, paramétrages, etc.), une approche déterministe, ne prenant pas en compte explicitement ces incertitudes, ne permet pas d'estimer celles associées aux débits prévus.

Une thèse actuellement en cours a pour objectif de mettre en place une chaîne de prévision des crues ensembliste en utilisant les prévisions ensemblistes de pluie AROME produites par Météo-France. Cette approche devrait permettre de mieux quantifier les incertitudes affectant les prévisions de débit.

Le travail est réalisé sur le bassin de l'Odet dans le Finistère, surveillé par le Service de Prévision des Crues « Vilaine et Côtières Bretons ». L'Odet est une rivière côtière. La ville de Quimper est régulièrement touchée par des inondations dues à la concomitance de pluies importantes sur le bassin versant et de marées hautes amplifiées par des surcotes d'origine maritime.

Le calculateur Datarmor a été utilisé en 2018 pour réaliser une étude d'incertitude globale de deux modèles hydrologiques : le modèle GRP, développé par IRSTEA, et le modèle MORDORTS, développé par EDF-DTG. Ces modèles serviront ensuite à alimenter un modèle hydraulique de prévision des crues (Mascaret). L'étude d'incertitude permet de déterminer quels sont les paramètres qui conditionnent le plus le débit simulé par le modèle, et permet donc d'orienter la stratégie de correction des débits simulés par assimilation de données.

Pour mener cette étude, il s'agit dans un premier temps de déterminer pour chacun des paramètres incertains du modèle (paramètres de calage, forçages météorologiques) une loi statistique le représentant. Ensuite, des simulations sont réalisées par tirage aléatoire des valeurs de ces paramètres incertains dans le but de calculer des indices de Sobol. Ces indices permettent de mesurer l'influence des différents facteurs d'entrée indépendants et leurs interactions sur la variable d'intérêt, le débit modélisé par le modèle hydrologique dans notre cas. Les indices de Sobol sont compris entre 0 et 1. Plus l'indice est élevé, plus le facteur d'entrée a de l'influence sur la variable de sortie du modèle.

Un grand nombre de simulations est nécessaire pour atteindre la convergence des indices de Sobol. Nous présentons ci-après quelques résultats obtenus pour le modèle GRP.

#### 6.1.16.2 Étude de sensibilité du modèle GRP

Pour mener cette étude, nous avons à disposition :

- des prévisions d'ensemble de pluie PEAROME. Ces prévisions d'ensemble ont 12 membres ;
- le modèle GRP qui possède 3 paramètres de calage. Pour cette étude, le modèle GRP a été utilisé sans l'assimilation de données.

Les paramètres incertains considérés pour l'étude de sensibilité sont les 3 paramètres de calage du modèle GRP ainsi que la pluie qui force le modèle. Pour cette étude 50000 runs ont été lancés. La figure 6-18 présente les indices de Sobol en fonction du temps pour un événement de crue du mois de janvier 2016.

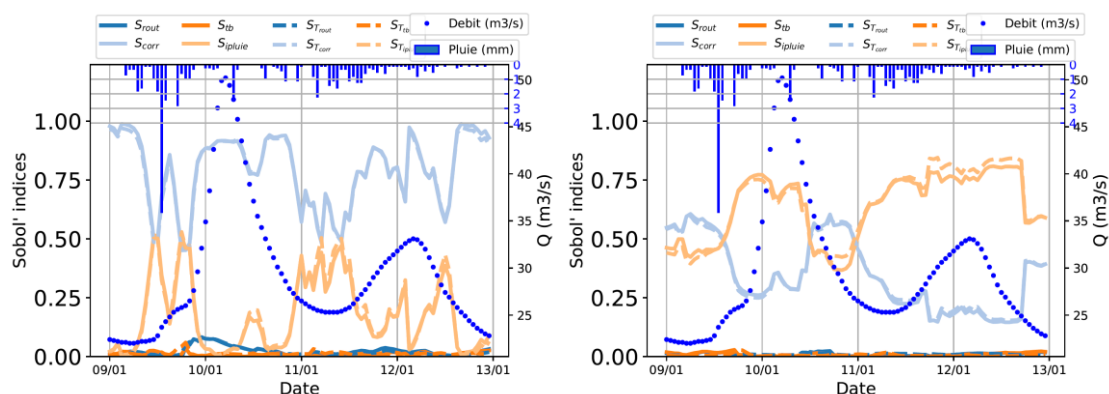


Figure 6-18 : Indices de Sobol avec le modèle GRP-ss-assim – Station Tréodet – Echéances 6 heures (gauche) et 21 heures (droite)

L'analyse de ces graphiques met en évidence les paramètres prépondérants en fonction de l'événement pluvieux. À longue échéance (21 heures), la pluie prend une importance croissante. En effet, plus l'échéance est importante, plus la quantité d'eau entrée dans le modèle va être différente selon le choix de la pluie, donc l'influence sur les débits simulés est plus grande. Nous mettons également en évidence le fait que le paramètre Corr est prédominant en dehors des événements pluvieux. Ce paramètre conditionne en partie le volume d'eau entrant dans le modèle.

Ces travaux seront poursuivis en 2019 par une analyse de sensibilité du modèle Mascaret sur l'Odet.

### 6.1.17 Étude de données de vent en Europe pour caractériser le potentiel de l'éolien en mer

*Léa Thiebaud, Responsable d'étude énergies au Cerema Eau mer et fleuves*

*Anne-Laure Tiberi-Wadier, Responsable d'étude modélisation au Cerema Eau mer et fleuves*

En 2016, le Commissariat Général au Développement Durable du Ministère de la Transition Écologique et Solidaire a demandé au Cerema d'effectuer une analyse de données de vent en mer, au large de pays européens afin de caractériser le potentiel de l'éolien en mer.

Une étude a donc été menée en 2017 et 2018. Des données issues de modèles météorologiques et de mesures in situ ont ainsi été collectées puis analysées pour :

- identifier les écarts entre modèles et mesures ;
- caractériser des variations spatiales et temporelles du vent ;
- identifier et comparer des formules pour l'extrapolation verticale du vent ;
- déterminer un facteur de charge théorique pour les éoliennes en mer.

### 6.1.17.1 Données étudiées

Les données étudiées proviennent de ré analyses météorologiques et de mesures in situ réalisées par des anémomètres fixés sur des mâts. La zone d'étude correspond à la partie nord-ouest de l'Europe. Les données étudiées sont les ré analyses CFSR, Era Interim à 10 mètres et des données issus des mâts de mesures et de bouées en mer. Les vitesses de vent ont été comparées pour certains évènements extrêmes et en moyenne.

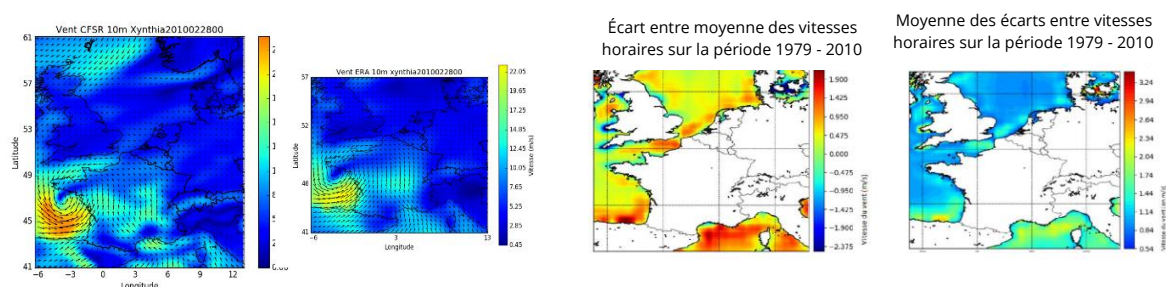


Figure 6-19 : Comparaison des données CFSR et Era Interim sur la période 1979 - 2010

Les jeux de données étudiés sont globalement cohérents. Les écarts les plus importants se situent à proximité des côtes et diminuent au large, là où les vitesses sont plus fortes.

### 6.1.17.2 Extrapolation verticale du vent

Les résultats de l'extrapolation verticale du vent des données CFSR à l'aide de plusieurs formules ont été comparés à des mesures in situ à différentes hauteurs.

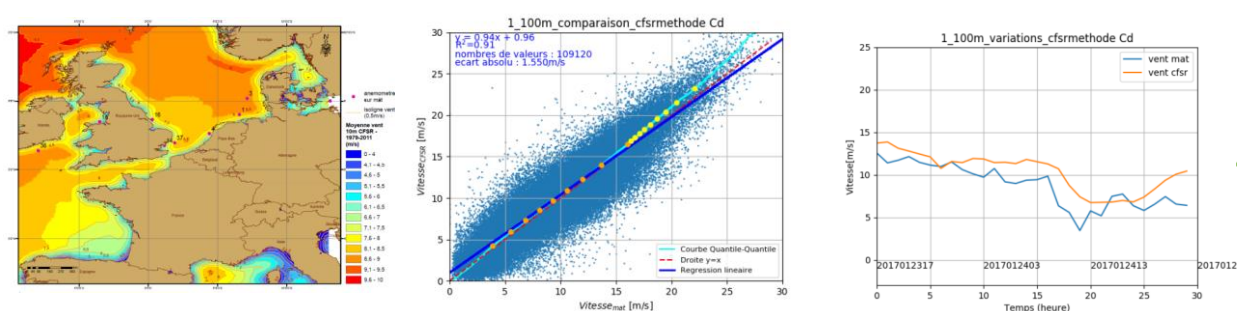


Figure 6-20 : Comparaison des vitesses issues de ré analyses CFSR extrapolées à 100 mètres par une loi logarithmique et des données du mât FINO 1 (point 1 sur la carte)

### 6.1.17.3 Variations du vent

Les variations du vent ont ensuite été étudiées à partir des données CFSR. Les moyennes mensuelles et horaires du vent ont été représentées sur des cycles annuels et journaliers de manière à identifier les périodes de vents forts et faibles bien que le régime de vent présente également des variations interannuelles.

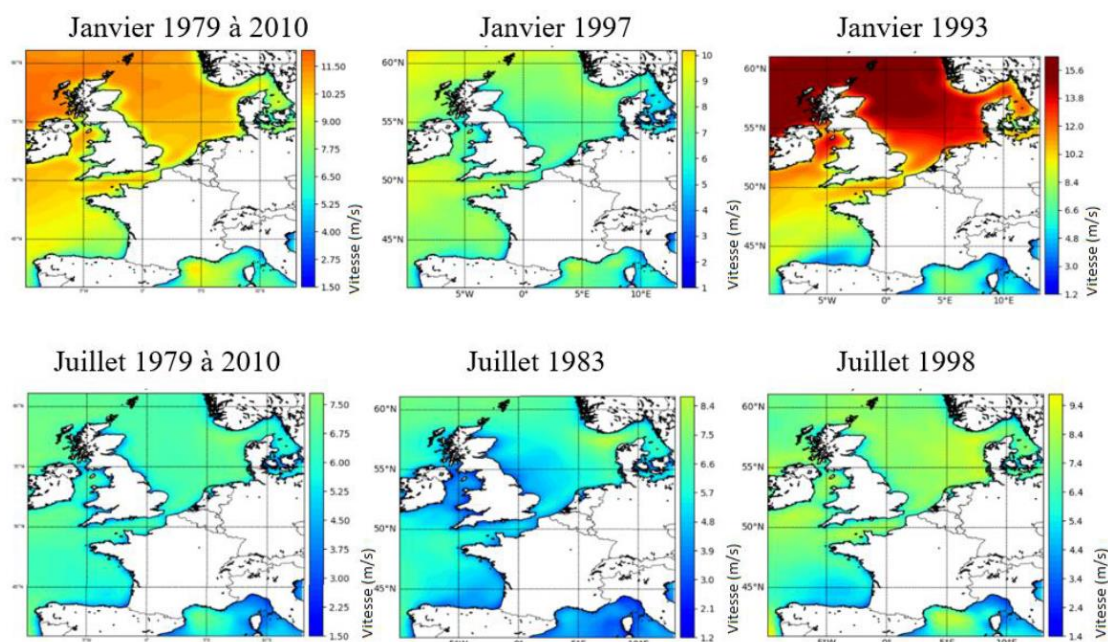


Figure 6-21 : Variations mensuelles du vent (données CFSR) sur la zone d'étude

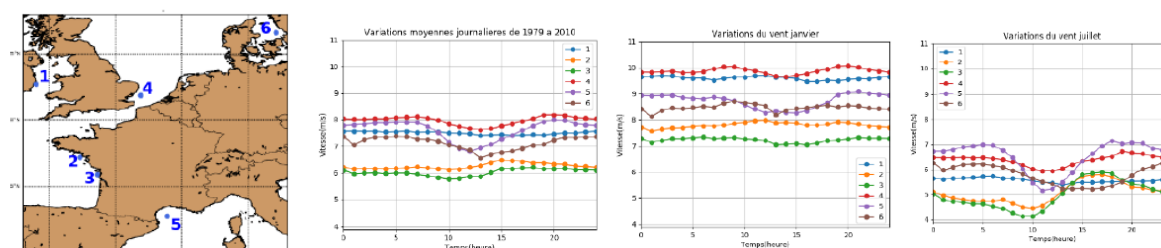


Figure 6-22 : Variations horaires du vent (données CFSR) en quelques points

Les données moyennées par heure permettent de visualiser des variations journalières plus marquées en été qu'en hiver notamment pour les points 2, 3 et 5 situés au sud de la zone d'étude. En période estivale les phénomènes de brises de mer, qui se traduisent par une augmentation de la vitesse du vent en journée, peuvent être plus marqués en ces points. Par ailleurs, les vents d'ouest dominants en Europe aux latitudes de la zone d'étude (LAUGEL, 2013) peuvent expliquer le décalage temporel des extremums du vent.

#### 6.1.17.4 Calcul du potentiel éolien

La production est évaluée pour chaque heure afin de déterminer le facteur de charge moyen de 1979 à 2010. Les vitesses de vent prises en compte sont issues des réanalyses du modèle CFSR extrapolées verticalement par une loi logarithmique puis moyennées sur la surface du disque balayée par les pâles de l'éolienne.

Le facteur de charge estimé peut atteindre 60 % de temps sur quelques secteurs de la mer du Nord. Cependant, il ne prend pas en compte :



- les variations du coefficient de puissance avec la vitesse de rotation des pâles ;
- l'extrapolation des vitesses et des directions du vent au-dessus de la couche limite atmosphérique ;
- la mise en arrêt des éoliennes liée à la maintenance des machines ;
- les pertes de production d'énergie liées aux effets de sillage au sein du parc, aux pertes électriques sur le réseau.

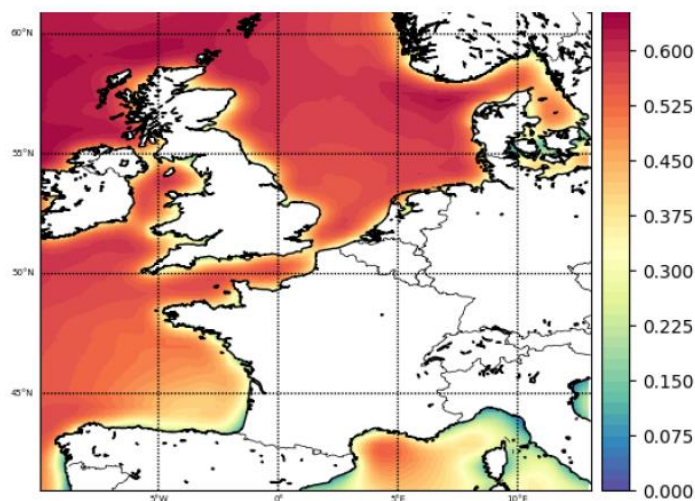


Figure 6-23 : Représentation du facteur de charge théorique moyen de 1979 à 2010

L'utilisation d'une courbe de puissance théorique aura tendance à surestimer le potentiel (JOURDIER, 2015). De plus, des données de vent plus précises pourraient permettre de mieux considérer des phénomènes de cisaillement, de turbulence ou des rafales qui peuvent impacter le fonctionnement des éoliennes.

#### 6.1.17.5 Utilisation de Datarmor

Datarmor a été utilisé pour l'extraction des re analyse au format NetCDF, l'extrapolation verticale des vitesses, les comparaisons entre modèles et mesures in situ, l'analyse des variations spatiales et temporelles du vent, l'extrapolation verticale du vent et le calcul de la production.

#### 6.1.18 Optimisation du frottement sur le fond en environnement macrotidal : application aux courants barotropes du Raz Blanchard

*Victor Quilfen (Shom), Cyril Lathuilière (Shom), Audrey Pasquet (Shom), Martial Boutet (Ifremer ODE/LOPS/OC)*

La mise en œuvre de fermes d'hydroliennes afin de tirer profit de l'énergie des courants de marée, nécessite une bonne connaissance de l'hydrodynamique locale et en particulier de la vitesse des courants de marée. En cela, les modèles numériques sont de précieux outils car ils permettent une représentation spatiale et temporelle de cette dynamique. A l'heure

actuelle, on estime qu'une erreur de 10 % sur le calcul de la vitesse des courants engendre une erreur de 30 % sur la prédiction du productible électrique, il est donc impératif de diminuer le plus possible cette erreur. Dans le cadre du projet Hydrodynamique du Raz Blanchard : Mesure et Modélisation (HYD2M) piloté par France Energies Marines (FEM) et l'Université de Caen, et dont le Shom est partenaire, nous nous sommes intéressés à la zone du Raz Blanchard, étroit bras de mer entre l'île d'Aurigny (*Alderney* en anglais) et le Cap de la Hague où les courants de marée peuvent atteindre 5 m/s, en faisant un endroit clé pour la mise en place d'hydroliennes. Le but de notre étude est d'optimiser le frottement sur le fond dans le modèle opérationnel du Shom HYCOM (HYbrid Coordinate Ocean Model) (e.g., Bleck, 2002 ; Chassignet et al., 2007) dans la zone du Golfe Normand-Breton, incluant le Raz Blanchard. La distance entre les frontières est et ouest est de 180 km. La distance entre les frontières nord et sud est de 188 km. Les données bathymétriques, provenant du MNT à 100 m de résolution issu du projet HOMONIM conjoint au Shom et à Météo France, sont interpolées sur la grille de calcul d'une résolution de 400 m pour 513\*498 éléments. Seuls les courants barotropes étant étudiés, le modèle ne dispose que d'une seule couche verticale, et les processus physiques tels que la propagation et la dissipation par déferlement des vagues ou encore les courants géostrophiques sont laissés de côté. La marée est forcée aux frontières ouvertes par l'atlas FES2014 (Carrère et al., 2014). Le modèle tourne sur 7 cœurs pour un nombre maximum de processeurs égal à 191. Les simulations sont lancées sous forme de jobs en « qsub » sur la queue mpi\_7 pour un « walltime » fixé à 04h30. Le temps de simulation pour 200 jours est de 03 heures. La validation du modèle est faite sur un jeu de données de la base de données du Shom (TDB) comprenant 33 marégraphes pour les hauteurs d'eau, et 3 ADCPs dans la zone du Raz Blanchard. Une première étude d'analyse et validation harmonique est réalisée afin d'obtenir la longueur de rugosité de fond ( $z_0$ ) optimale dans l'expression du coefficient de frottement ( $C_b = \kappa^2 / \left( \ln\left(\frac{h}{z_0}\right) - 1 \right)^2$ ) d'une loi quadratique pour les 21 composantes de marée. Nous obtenons une valeur optimale uniforme de 10 mm. La Figure 6-24 illustre le résultat obtenu pour l'onde M2 en amplitude (a) et en phase (b).

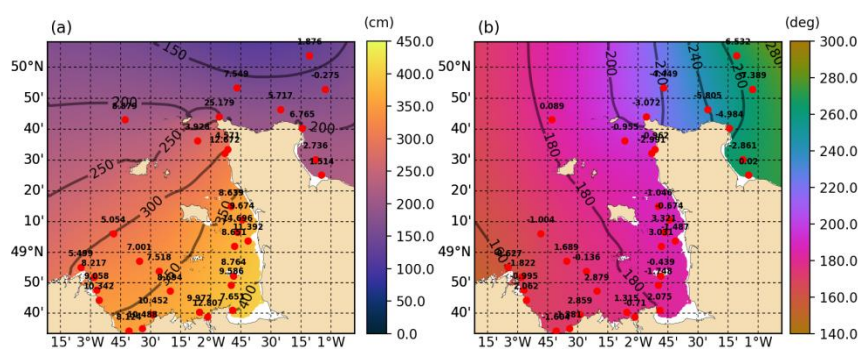


Figure 6-24. Cartes des amplitudes (a) et phases (b) modélisées pour un  $z_0$  égal à 10 mm. Les points en rouge indiquent l'écart relatif à la mesure marégraphique (en cm).

Une deuxième étude porte sur la création de cartes optimisées de  $z_0$ , suivant deux méthodes :

Sediment classes	$z_0$ (mm)
Pebbles to Rocks	22.55
Gravels to Pebbles	16.42
Coarse Sands to Gravels	6.02
Mud to Coarse Sands	8.83

Tab. 1. Valeurs du  $z_0$  associées à chaque classe de sédiment.

**Méthode 1 :** En calquant des valeurs préalablement optimisées du  $z_0$  dans une maquette globale du modèle HYCOM de la zone Manche-Gascogne (Boutet, 2015) sur la distribution spatiale des grains sédimentaires dans le Golfe Normand-Breton (maquette régionale) en simplifiant le problème à 4 classes de sédiments (blocs de roche, cailloutis, graviers et sables).

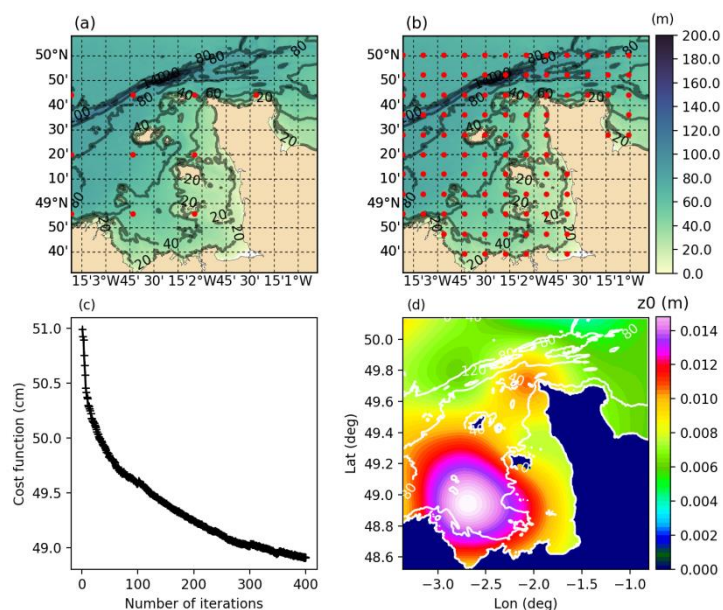


Figure 6-25 : Méthode d'optimisation du  $z_0$  par algorithme SPSA. La localisation des points de collocation est donnée en (a) pour 10 points et (b) pour 101 points. La réduction de la fonction de coût après chaque itération est donnée en (c) et la carte optimisée du  $z_0$  résultante est donnée en (d).

**Méthode 2 :** Grâce à l'utilisation d'un algorithme *Simultaneous Perturbation Stochastic Approximation* (SPSA) (e.g., Spall, 1992 ; Boutet, 2015) dont l'objectif est de minimiser de manière itérative une fonction de coût basée sur l'écart-type des différences entre le niveau d'eau modélisé et mesuré à chaque marégraphe. A chaque itération le code stochastique estime le  $z_0$  en un certain nombre de points de collocation progressivement augmentés de  $2 \times 2$  à  $10 \times 10$  (Fig. 2a,b), et ce pour 4 calculs en parallèle (4 jobs Datarmor en parallèle lancés sur 7 cœurs) représentant 4 perturbations différentes du système. Le modèle tourne sur 100 jours, et chaque simulation/itération dure 01h30. L'algorithme d'optimisation est stoppé après 400 itérations, lorsque la valeur de la fonction de coût n'évolue quasiment plus (Fig. 2c). Les valeurs de  $z_0$  obtenues aux points de collocations sont ensuite interpolées sur l'ensemble de la grille (Fig. 2d).

Enfin, le modèle est testé sur 100 jours et les données prédites sont comparées aux données mesurées. L'allure des champs de courants dans le Raz Blanchard est illustrée à la Figure 6-26 Les résultats aux 3 ADCPs (Figure 6-27a) sont comparés en terme de RMSE sur le courant total (Tab. 2) et de MAPE sur deux classes de vitesses (« faibles » < 2m/s et « fortes » > 2m/s) (Figure 6-27b) pour les quatre paramétrisations du coefficient de frottement suivantes : (i)  $C_B = 2.5 \times 10^{-3}$  ; (ii) loi log avec  $z_0 = 10$  mm obtenu par analyse harmonique ; (iii) loi log avec  $z_0(d_{50})$  variable obtenu par la Méthode 1 ; (iv) loi log avec  $z_0(x,y)$  obtenu par la Méthode 2.

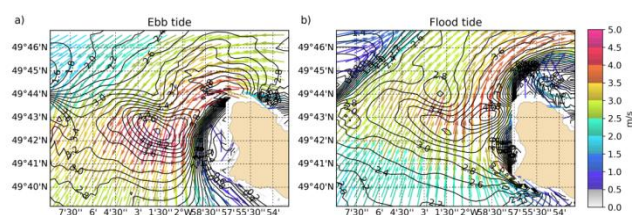
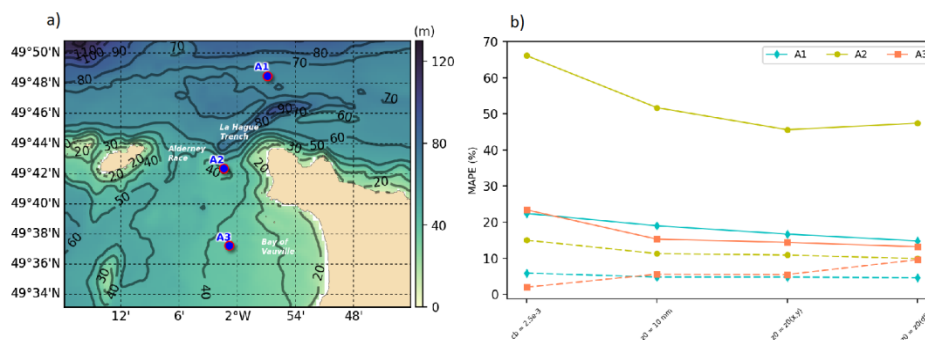


Figure 6-26 : Champs de courants prédits par HYCOM dans le Raz Blanchard au jusant (a) et au flot (b).

$C_B$ law	A1	A2	A3
$C_B = 0.0025$	0.17	0.54	0.23
$C_B = C_B(h, z_0)$ with $z_0 = 10$ mm	0.15	0.39	0.14
$C_B = C_B(h, z_0)$ with $z_0 = z_0(x, y)$	0.14	0.37	0.14
$C_B = C_B(h, z_0)$ with $z_0 = z_0(d_{50})$	0.12	0.34	0.13

Tab. 2. RMSE (m/s) en trois points du Raz Blanchard pour différentes paramétrisations du coefficient de frottement.



Les résultats montrent qu'une loi logarithmique du  $C_b$  fonction à la fois de la hauteur d'eau et de la rugosité  $z_0$  diminue la RMSE de 2 cm/s à 15 cm/s selon le cas. Si en plus le  $z_0$  est variable en espace, la RMSE diminue de 1 cm/s à 5 cm/s selon le cas. Les vitesses élevées ( $> 2$  m/s) sont mieux reproduites que les vitesses faibles ( $< 2$  m/s). L'utilisation d'une méthode d'optimisation SPSA permet néanmoins de réduire l'erreur sur ces vitesses « faibles » d'environ 5 % dans le Raz Blanchard. Dans réalité, en plus d'être spatialement variable, le  $z_0$  est temporellement variable suivant l'approche des courants sur la bathymétrie accidentée du fond (différente du flot au jusant) et l'intensité des courants (forte en vives-eaux mais plus faible en mortes-eaux), d'autant plus dans un contexte topographique complexe tel que le Raz Blanchard. Une paramétrisation plus complète serait nécessaire pour reproduire ces effets temporels.

## 6.2 Entreprises

### 6.2.1 BV Solutions M&O : Cas test de slamming de maxi-trimaran de course au large

La Route du Rhum 2018 est une course au large en solitaire, reliant Saint Malo à la Guadeloupe. Parmi les classes de bateaux ayant couru la dernière édition, la classe Ultim 32/23 de maxi-trimarans a été particulièrement médiatisée, notamment du fait des avaries subies par 2 des 6 bateaux (Gitana, Banque Populaire) ayant entraîné leur abandon.

BV Solutions M&O travaille régulièrement pour des études dans le domaine de la voile de compétition, en lien direct avec les teams de course au large, les architectes et les designers de ces bateaux. Un cas-test de simulation numérique de slamming de maxi-trimaran dans la houle a donc été mis en place afin d'initier des discussions avec les teams possédant ces trimarans. En effet, le phénomène de slamming (impact d'un corps dans l'eau avec une cinématique rapide) peut engendrer des efforts très importants sur la structure des bateaux ; l'étude de ces cas permet de mieux dimensionner la structure.

Les objectifs de ce cas-test sont les suivants :

- Réaliser une simulation réaliste de slamming d'un trimaran dans les vagues (impact d'un flotteur dans la vague)
- Réaliser des visuels attractifs (images, vidéos) à montrer aux teams de course au large

#### 6.2.1.1 Mise en place du modèle numérique

Le solveur numérique utilisé pour cette simulation est SPH-flow, co-développé par Centrale Nantes et Nextflow Software. C'est un solveur particulière (méthode lagrangienne) qui ne nécessite qu'un maillage surfacique du corps étudié, le fluide étant modélisé par des particules en interaction les unes avec les autres. Différentes boîtes de raffinement sont appliquées autour des zones d'intérêt (flotteur, foil) afin de diminuer la taille des particules et de capter l'écoulement plus précisément.

La géométrie de ce trimaran fictif a été reconstruite par BV Solutions M&O.

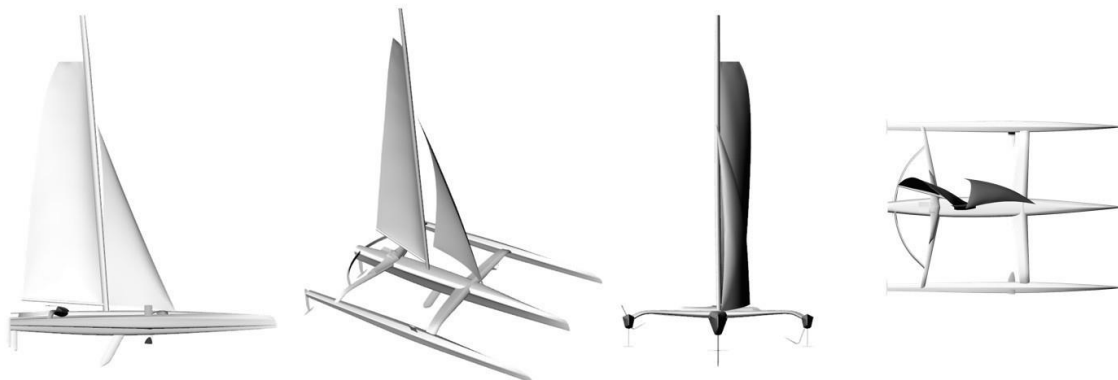
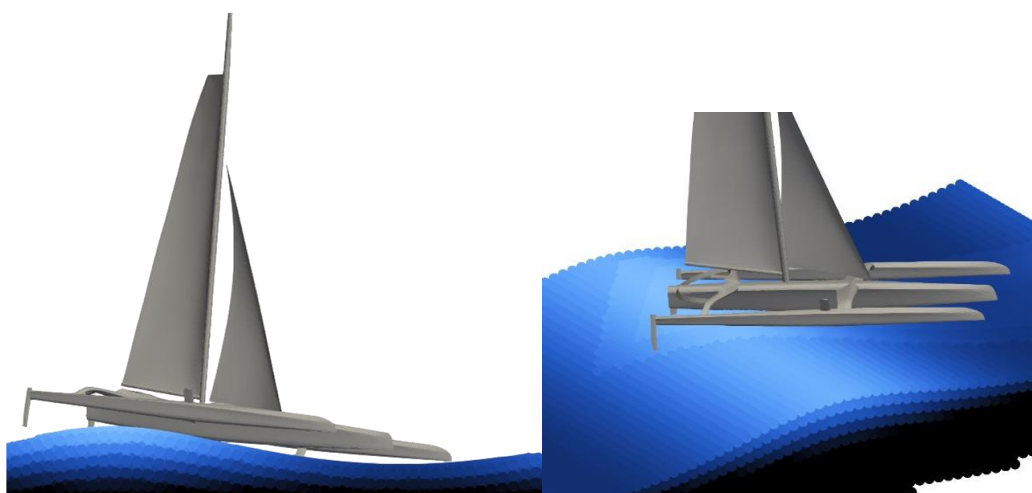


Figure 6-28 : Géométrie du maxi-trimaran

Les conditions de simulation ont été définies de manière à s'approcher le plus possible des conditions de navigation rencontrées par les trimarans lors des avaries de la Route du Rhum. Ainsi, un effort aérodynamique a été appliqué pour reproduire la force propulsive et le moment de gîte du vent dans les voiles. Le trimaran navigue dans le sens de la houle, à 150° de la direction de propagation de celle-ci. La houle régulière appliquée dans la simulation correspond à l'état de mer rencontré par les trimarans. Les paramètres sont les suivants

- 28 nœuds de vent apparent à 80° d'angle de vent apparent
- Vitesse bateau 35 nds
- Surface de voiles 326 m<sup>2</sup>
- Houle de hauteur 3.50 m, longueur d'onde 50.0 m



*Figure 6-29 : Positionnement du maxi-trimaran dans la houle*

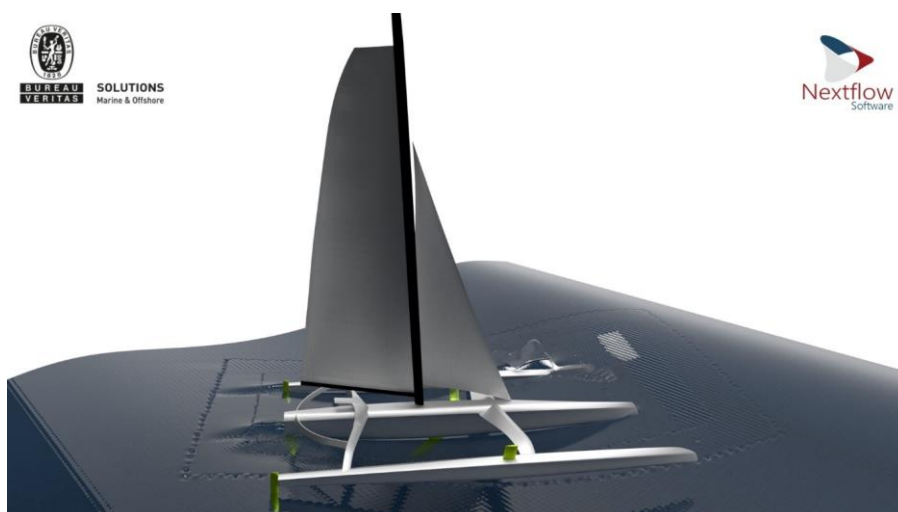
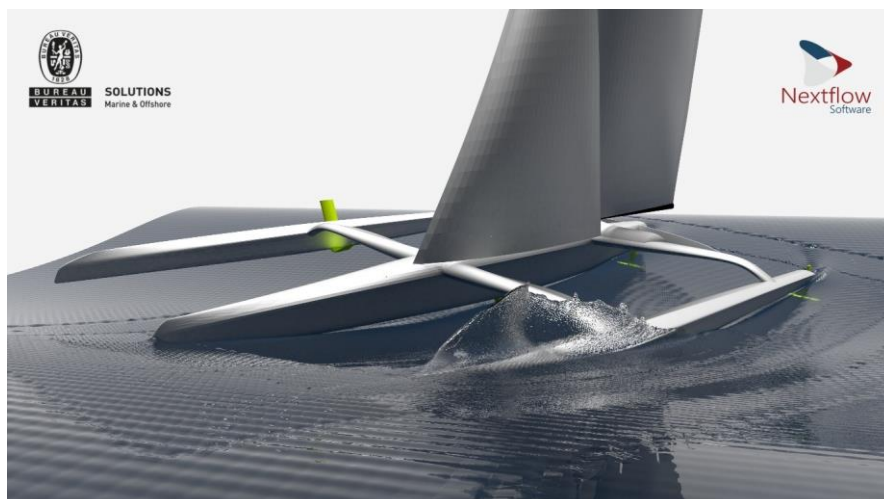
### 6.2.1.2 Résultats et rendus

Les résultats obtenus en termes d'efforts (forces et moments), de vitesse et d'accélération du trimaran impactant la houle ont permis de valider la cohérence du modèle.

L'un des enjeux étant la création de visuels attractifs pour de potentiels clients, les données volumiques ont été confiées à Nextflow Software afin de réaliser une vidéo en rendu blender (rendu réaliste avec indice de réfraction de l'eau, transparence, matériaux pour les corps, effets d'ombres et de lumières). Des images sont présentées ci-dessous.

### 6.2.1.3 Conclusions

Ce cas-test de slamming de maxi-trimaran dans la houle a été réalisé par BV Solutions M&O en faisant les simulations numériques sur le cluster Datarmor (environ 10 calculs de 10h sur 224 CPU). Nextflow Software a réalisé les post-traitements en rendu blender sur Datarmor également (environ 10 h sur 56 CPU). Suite à ce cas-test, BV Solutions M&O a commencé à travailler avec le cabinet de calcul de structure GSea Design pour des études sur les futurs trimarans Macif et Banque Populaire.



*Figure 6-30 : Rendus blender de la simulation de slamming du trimaran*

## 6.2.2 BV Solutions M&O – RealTide Project

Le projet RealTide est un projet Européen impliquant de nombreux partenaires qui a pour objectif d’approfondir la maîtrise et la connaissance autour de l’exploitation de l’énergie des courants marins. Il traite des différentes problématiques rencontrées au cours du processus de conversion d’énergie, de la compréhension du courant incident à la distribution de l’énergie collectée sur le réseau terrestre.

Le libellé du projet est le suivant: “Advanced monitoring, simulation and control of tidal devices in unsteady, highly turbulent realistic tide environments — RealTide”

BV Solutions M&O est responsable du développement d’un modèle numérique permettant l’étude des performances hydrodynamiques d’hydroliennes dans des conditions environnementales réalistes (prise en compte de la turbulence, de l’influence des vagues...). Ce modèle est développé en interaction directe avec l’Université d’Edinburgh qui conduit notamment une campagne de mesures sur site pour fournir les données d’entrées nécessaires.

L'étude hydrodynamique d'hydroliennes conduite dans le cadre du projet se concentre en particulier sur 4 points:

- Etude hydrodynamique de la turbine en conditions idéales
- Simulation d'un courant de marée incident réaliste (turbulence, profil de vitesse, vagues...)
- Interaction de la turbine avec la chaîne de conversion d'énergie (modèle « tide-to-wire »)

### 6.2.2.1 Etude de performance et calcul des efforts hydrodynamiques

Une première phase du projet conduite dans le courant de l'année 2018 a permis de déterminer les courbes de performances de l'hydrolienne D12 de Sabella. Les calculs ont été réalisés à l'échelle réelle mais aussi à l'échelle maquette pour fournir une base de comparaison directe aux résultats obtenus en essais bassins. Cela s'est traduit par la mise en place d'un modèle de calcul simplifié ne prenant pas en compte la complexité du courant dans un premier temps. Différents tests de sensibilité sur le maillage, le pas de temps, ou la convergence des résultats ont été conduits. L'hydrolienne D10 de Sabella sur laquelle de nombreux résultats expérimentaux étaient disponibles a servi de référence pour vérifier la juste calibration du modèle. Finalement, les efforts hydrodynamiques calculés, nécessaires à l'étude structurale de la pôle de l'hydrolienne, ont été fournis à Sabella (champs de pression, efforts moyens...).

En parallèle la même étude est actuellement conduite sur une hydrolienne générique 3 pâles fournie par l'Université d'Edimbourg qui servira de deuxième référence au cours du projet. Cela permettra de couvrir une plus large gamme de types d'hydroliennes.

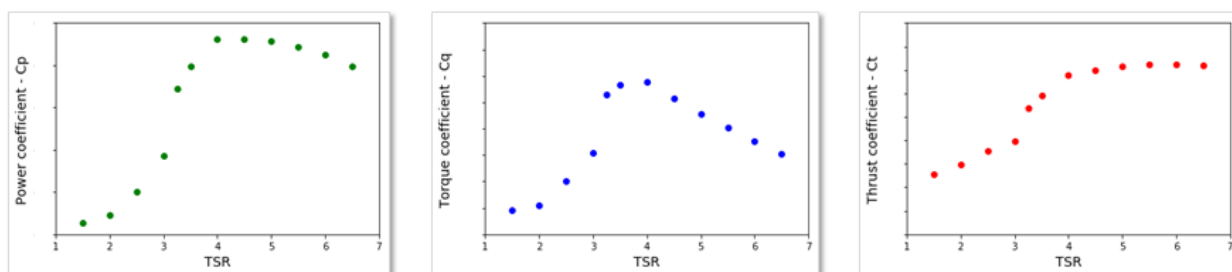


Figure6-31 :: Performance curves on D12 Sabella tidal turbine as a function of TSR

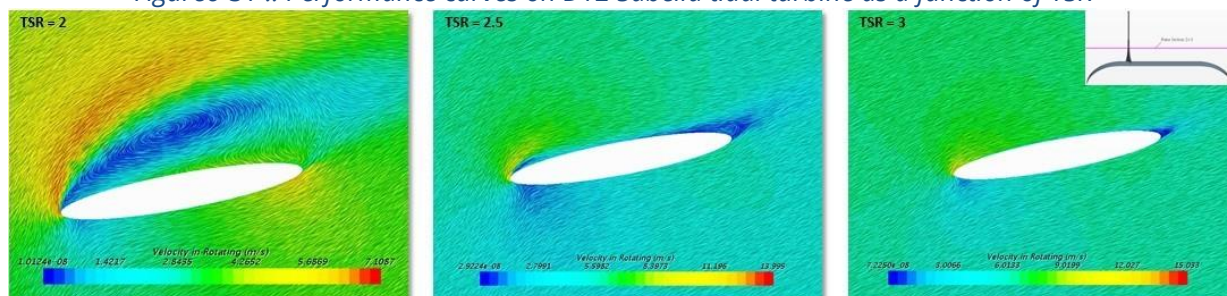


Figure6-32 : Velocity profile around the blade at TSR = 2, TSR = 2.5 and TSR = 3 for a uniform unidirectional inflow velocity  $U_{current}=2.75$  m/s



### 6.2.2.2 Modélisation de conditions environnementales réalistes

Suite à cette première étude hydrodynamique, une seconde phase du projet commencée en début d'année 2019 vise à intégrer des conditions environnementales réalistes dans le modèle numérique. Ceux-ci ont un effet direct tant sur les contraintes subies par l'hydrolienne que sur le sillage généré en aval. Les principales composantes de courant à intégrer sont les suivantes : turbulence, profil vertical de vitesse, interaction houle-courant. Une première étape de revue bibliographique a eu pour objectif de recenser les méthodes numériques les plus à même de modéliser ces paramètres avec précision.

Suite à celle-ci, un premier travail sur le maillage visant à prendre en compte un profil vertical de vitesse ainsi qu'une zone de sillage bien résolue a été effectué. Cela a permis de définir un nouveau modèle plus à même de rendre compte de l'influence des facteurs environnementaux. Dans un second temps, et en collaboration directe avec Sabella, des cas de fonctionnement réalistes de l'hydrolienne ont été identifiés et sont actuellement modélisés. Ces cas incluent par exemple des situations de fonctionnement critique (roue libre, arrêt d'urgence, conditions extrêmes annuelles), mais aussi des cas rencontrés au quotidien tels qu'un courant avec un angle d'incidence non nul, un couplage houle-courant... L'étude de ces scénarii fournira des données qui permettront l'identification des facteurs environnementaux clés à modéliser, de guider les choix de design autour de l'hydrolienne et d'estimer les capacités de conversion disponibles selon les cas de fonctionnement.

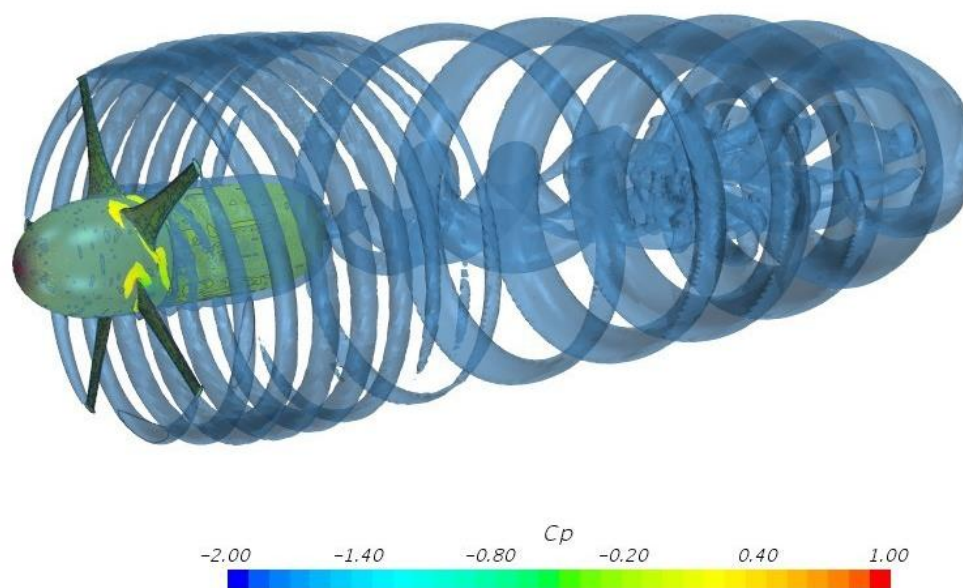


Figure6-33 :  $C_p$  on D12 surface with iso-surface of Q-Criterion of 0.05 s-2 in domain

L'intégration de la turbulence notamment est un point particulièrement sensible du fait de sa forte influence sur l'écoulement. Un test de sensibilité a été réalisé sur les paramètres physiques associés (longueur caractéristique turbulente, intensité turbulente, ratio de viscosité turbulente) de façon à quantifier leur influence sur l'écoulement et sur la performance de l'hydrolienne. Ces calculs ont permis de rendre compte des limites des simulations de type RANSE (Reynolds Averaged Navier-Stokes Equations) qui tend à lisser les effets de la turbulence. Il est alors nécessaire d'utiliser des modèles numériques plus fins tels que la DES (Detached Eddy Simulation) mais qui génèrent un temps de calcul beaucoup plus

long. Ces méthodes numériques seront étudiées plus en détail dans le courant de l'année à travers une série de simulations qui permettra de comparer les résultats RANSE et DES.

La suite du projet permettra d'aboutir à un modèle complet capable de rendre compte avec précision de l'influence des facteurs environnementaux sur le fonctionnement de l'hydrolienne. Un couplage avec un module de conversion d'énergie optimisant la rotation de la turbine sera ajouté dans un troisième temps. Utilisé en parallèle de mesures sur site et d'essais en bassin, ce modèle de simulation numérique pourra être utilisé aussi bien pour fournir des données nécessaires à l'étude structurelle de nouvelles turbines que pour étudier leur capacité de conversion d'énergie en conditions réelles ou pour optimiser la disposition des hydroliennes en ferme.

### 6.2.3 BV Solutions M&O : Optiroutes



## OPTIROUTES

Optimisation des performances, de la sécurité et de la durée de vie des navires en conditions réelles.

Application au routage et à l'optimisation des carènes

OPTIROUTES est un projet FUI de 3 ans qui vise à réduire la consommation énergétique des navires, à améliorer leur sécurité et à augmenter leur durée de vie en intégrant dans les logiciels de routage une prise en compte précise de la réponse du navire à l'environnement dans lequel il évolue (état de mer, vent ...).

Pour cela, le projet consiste à mettre en place un outil de routage capable de prendre en compte le comportement hydrodynamique, aérodynamique et hydro / structure du navire.

Les développements réalisés concernent non seulement le passage d'un modèle océano-météo global à un modèle déterministe représentatif pour l'étude des performances du navire sur un état de mer donné, mais aussi la réduction de modèle pour définir en un minimum de calculs les performances globales d'un navire, ou encore des algorithmes de routage innovant intégrant l'ensemble des objectifs et contraintes d'un routage multi-objectif.

A titre d'exemple, il sera possible de définir la route optimale d'un navire en minimisant sa consommation, avec la contrainte d'arriver au point d'arrivée à une heure précise, en tirant partie des effets du vent, tout en évitant des zones pour lesquelles la sécurité peut être critique (roulis paramétrique, impact de slamming ou de green-water ...).

Les travaux issus de ce projet permettront aussi d'améliorer les performances et la sécurité des navires dès le stade de la conception, dans un environnement réel incluant les vagues et le vent.

Le projet OPTIROUTES est labellisé par le Pôle Mer et le Pôle EMC2.

En 2018, l'utilisation de DATARMOR a permis de poursuivre les développements du modèle navire sur la prédiction de la résistance mer calme (effets des tunnels d'étrave, de la profondeur d'eau) et des effets du chargement sur les efforts aérodynamiques. Les paragraphes ci-dessous synthétisent les résultats obtenus.

### 6.2.3.1 Quantification de la résistance ajoutée par les tunnels d'étrave

Afin de quantifier la résistance ajoutée due aux tunnels d'étrave, 32 calculs CFD et 18 000 h CPU ont été nécessaires. La résistance à l'avancement est légèrement augmentée par la présence des tunnels d'étrave (+1.38% en moyenne). Pourtant, la résistance visqueuse est légèrement diminuée par la présence des tunnels grâce à une zone de décollement à l'arrière du second tunnel (visible sur les images ci-après). L'augmentation de résistance est donc due à la résistance de pression qui augmente à cause d'une zone de surpression à l'intérieur des tunnels : le fluide vient taper contre le chanfrein des tunnels.

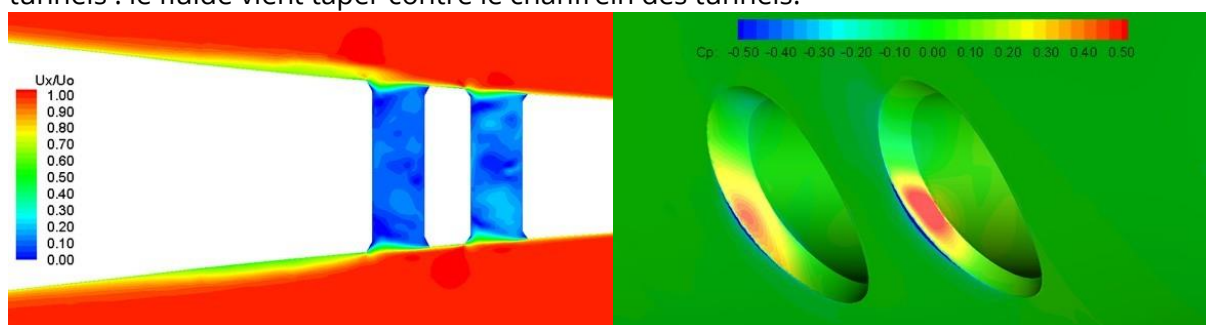


Figure 6-34 : Zone de décollement et zone de surpression induites par la présence des tunnels

L'effet des tunnels d'étrave sur la résistance à l'avancement diminue quand le tirant d'eau augmente car la résistance totale augmente avec le tirant d'eau.

### 6.2.3.2 Effet de la profondeur d'eau sur la résistance à l'avancement

Afin d'étudier l'effet de la profondeur d'eau, 30 calculs CFD et 67 000 h CPU ont été nécessaires. Cette étude a montré que la profondeur d'eau sous le bateau influençait largement l'enfoncement dynamique du navire ainsi que sa résistance à l'avancement (+36% de résistance à l'avancement dans les cas les plus extrêmes). Elle a aussi montré qu'à partir de 50m d'eau sous le bateau, la résistance à l'avancement s'approchait de la résistance en profondeur infinie (+3% au maximum). En revanche, l'enfoncement dynamique présente des écarts bien plus importants : jusqu'à +600% et toujours +15% avec 100m d'eau sous le navire.

Les images suivantes montrent l'impact de la profondeur d'eau sur le champ de vague : quand la profondeur diminue, un creux se forme autour du navire et le navire est aspiré vers le fond (ce qu'on appelle le « squat effect »).

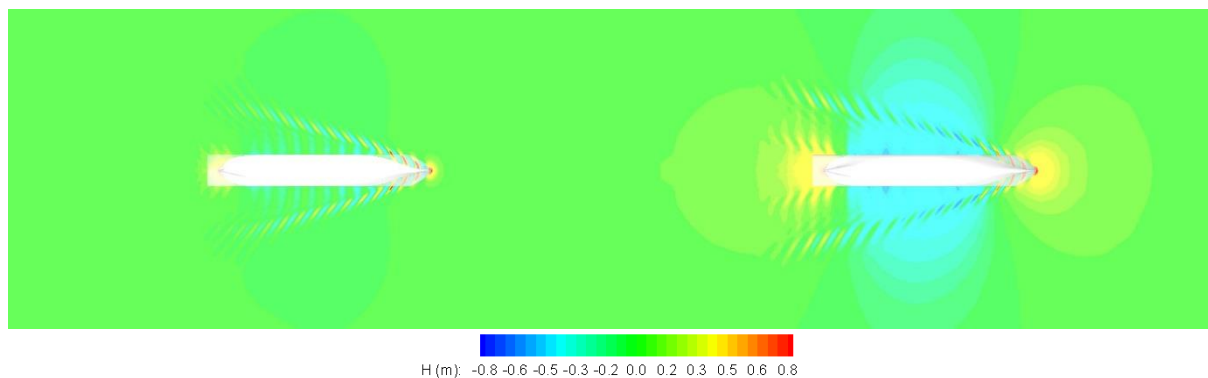


Figure6-35 : Champs de vague en profondeur infinie (gauche) et avec 5m d'eau sous le navire (droite)

### 6.2.3.3 Etude aérodynamique de chargements réels d'un porte-conteneurs

Une première étude sur des chargements théoriques (navire avec tous les conteneurs sur le pont, avec aucun conteneur sur le pont et avec une rangée de conteneurs sur deux) avait été menée en 2017. L'étude de ces cas théoriques avait montré une forte influence du chargement sur les coefficients aérodynamiques. Plusieurs cas de chargements réels ont depuis été évalués afin de savoir s'ils pouvaient se rapprocher d'un des 3 cas théoriques étudiés précédemment. Cette nouvelle étude a nécessité 180 000 h CPU. Elle montre que les chargements réels présentent des coefficients aérodynamiques situés entre le cas de chargement complet et le cas de chargement avec une rangée de conteneurs sur deux.

Les images suivantes montrent le coefficient de pression pour 6 cas de chargement différents.

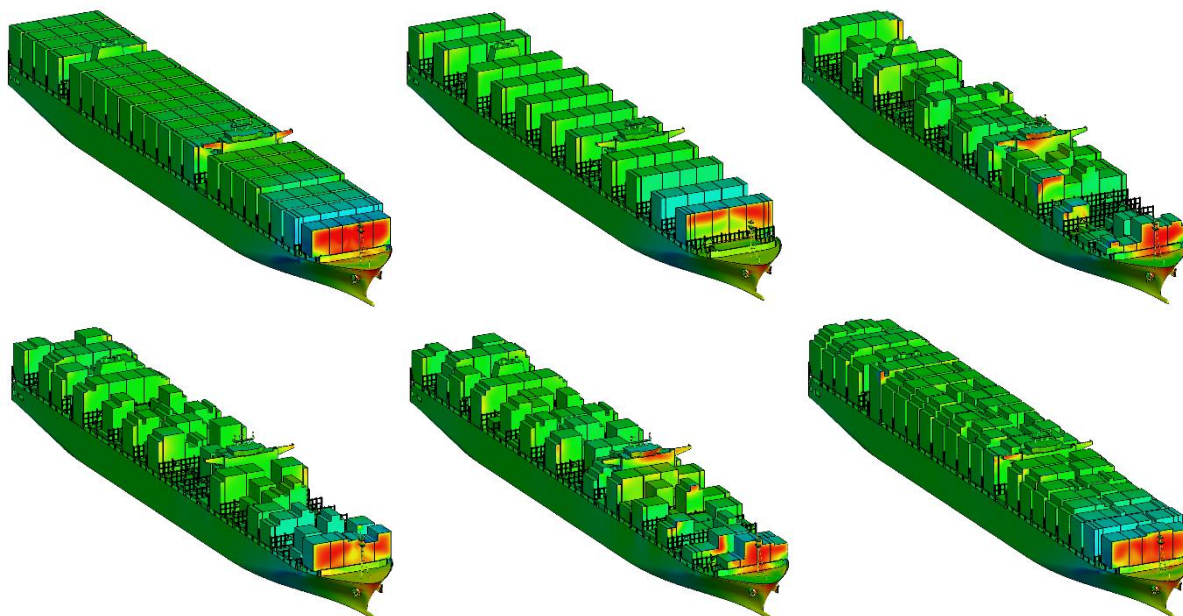


Figure6-36 : Coefficient de pression sur le navire pour 6 cas de chargement différents

## 6.2.4 EXWECS

*Extreme Weather Expertises* (EXWEXs) a pour spécialité la détermination des extrêmes météorologiques, océanographiques et climatiques sur des zones ou des sites pour lesquels il n'existe pas ou peu de mesures des événements. Pour ce faire, EXWEXs développe et met ainsi en œuvre des méthodes originales de détermination de ces extrêmes à l'aide d'algorithmes à la pointe des connaissances scientifiques en météorologie et en océanographie. L'utilisation combinée de données observées in-situ ou distantes (satellites) et de sorties de modèles numériques est une compétence importante de la société.

Ces traitements utilisant de larges ressources de calcul et génèrent de forts volumes de données nécessitant des infrastructures adaptées. Dans ce cadre, le cluster de calcul Datarmor permet à EXWEXs de déployer des configurations à haute résolution.

Un domaine d'expertise d'EXWEXs consiste en la modélisation et la prévision des systèmes convectifs, qui constituent une source de préoccupation majeure en termes de sûreté et de sécurité pour les clients de la société. Par leur caractère soudain et très localisé, les fronts de rafale associés à la convection violente sont aussi dangereux que difficiles à prévoir pour les modèles numériques de moyenne résolution.

Afin de répondre à cet enjeu, EXWEXs a développé un système de prévision atmosphérique intégré nommé ICFS (*Integrated Core Forecast System*). ICFS permet une modélisation fine de ces phénomènes convectifs notamment en mettant en œuvre des grilles de calcul à haute (quelques kilomètres) ou très haute résolution (quelques centaines de mètres). ICFS a notamment été utilisé pour la Coupe de l'America au sein d'*Emirates Team New Zealand* et a largement participé à la victoire du fait de la précision hectométrique de ses prévisions sur le vent et la turbulence sur un plan d'eau.

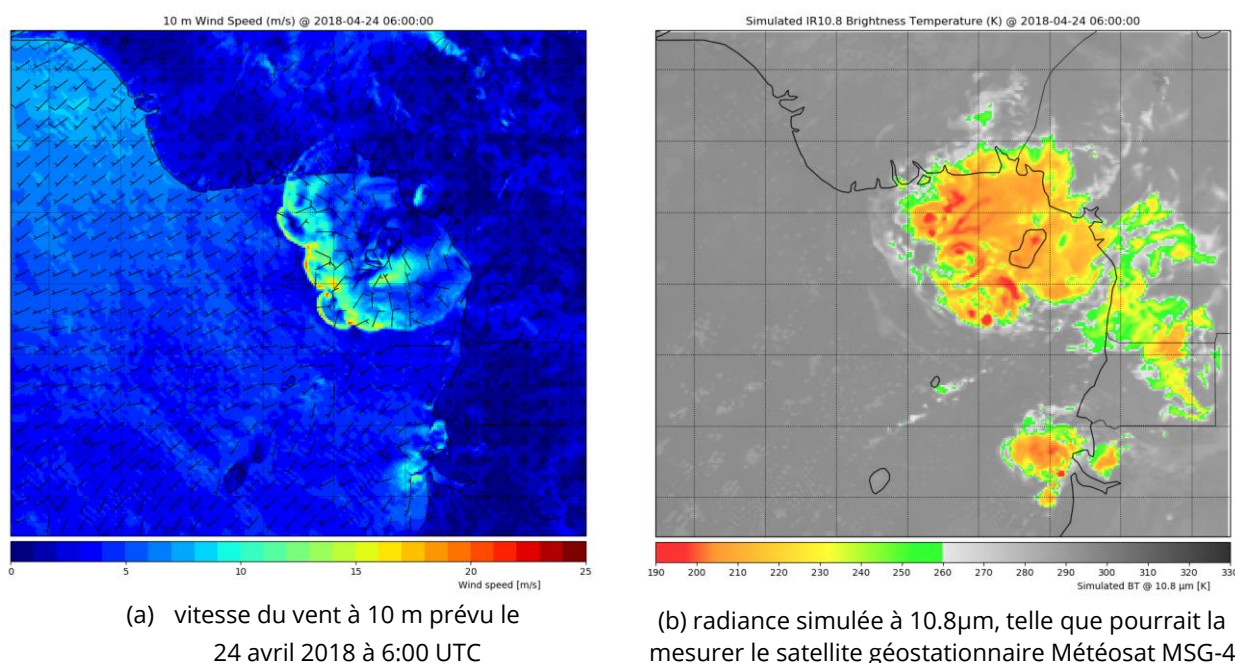


Figure 6-37 : Prédiction d'un système convectif au dessus du Golfe de Guinée par le modèle ICFS

La Figure 6-37 montre le résultat d'une prévision d'un système convectif sur une grille de maille 2 km, issue d'une configuration sur le Golfe de Guinée mise en place dans le cadre d'un système opérationnel de prévision déployé sur Datarmor au printemps 2018. Cette région étant en effet une préoccupation principale pour de nombreux opérateurs, civiles et militaires.

## 6.3 Applications opérationnelles

Il s'agit de traitements de données, de modèles qui tournent de manière automatique et quotidiennes sur le PCDM afin de produire des jeux de données pour la communauté scientifique.

### 6.3.1 Centre Aval de Traitement des Données SMOS – CATDS

SMOS (Soil Moisture and Ocean Salinity) est un satellite du programme commun d'observation de la Terre de l'ESA, du CNES et du CDTI, dédié à la mesure de l'humidité des sols (Soil Moisture - SM) et la salinité des océans (Ocean Salinity - OS). Il a été lancé fin 2009 et est pleinement opérationnel depuis mai 2010.

Chaque jour, le CATDS reçoit les données de température de brillance mesurées par le satellite SMOS.

Plusieurs traitements permettent :

- D'agréger à la journée les fauchées de température de brillance,
- D'inverser les données pour obtenir l'humidité des sols et la salinité de surface des océans,
- De filtrer et d'agréger ces données pour obtenir des cartes globales décadaires et mensuelles,

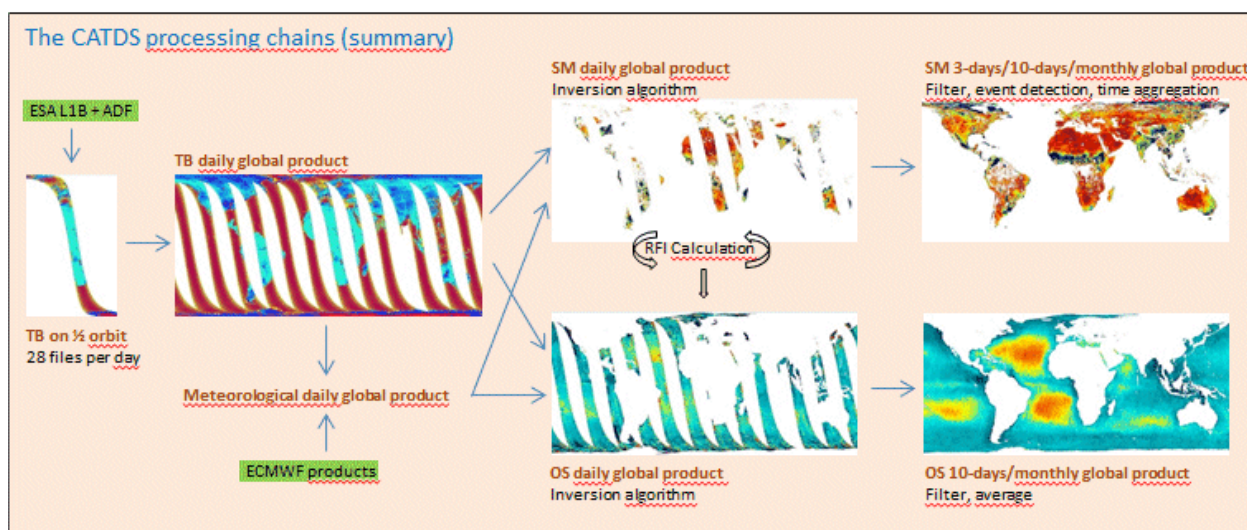


Figure 6-38 : Les chaînes de traitement opérationnelles du CATDS (résumé)

L'ensemble de ces traitements sont exécutés sur le serveur de calcul Datarmor et en utilisant ses capacités de stockage.

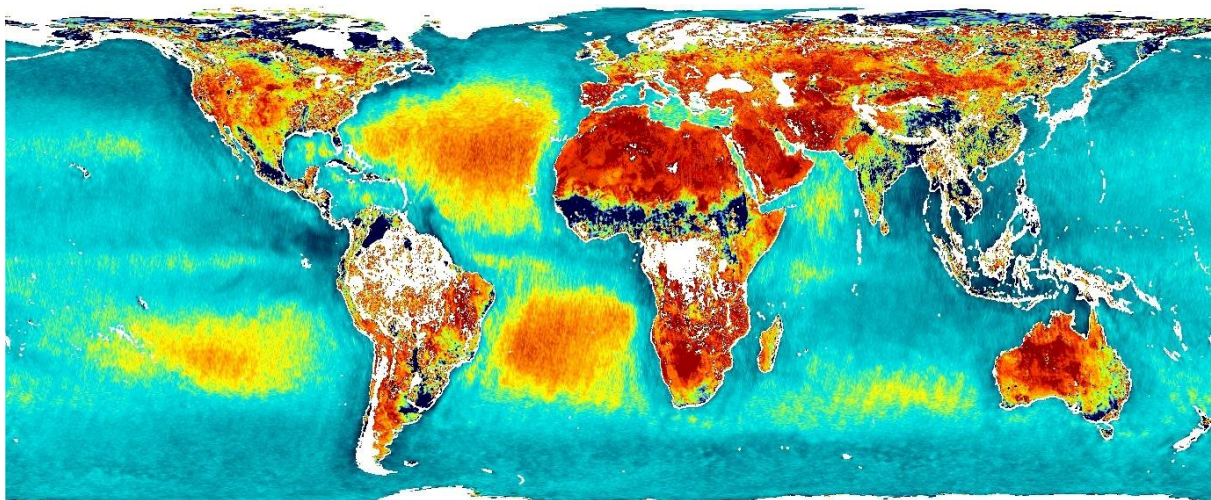


Figure 6-39 : Carte mensuelle couplée humidité des sols et salinité des océans (à partir des produits numériques du CATDS)

En 2018, Datarmor a principalement été utilisé par le CATDS pour la production opérationnelle quotidienne.

### 6.3.2 MARC

Le projet MARC (Modélisation et Analyse pour la Recherche Côtière) a débuté en 2016 suite au projet d'océanographie côtière opérationnelle PREVIMER. Il alimente les jeux de données des modèles hydrodynamiques MARS 2D, MARS3D, éco-MARS3D, des modèles d'états de mer WW3 et les données satellitaires de recherche développés dans le cadre de PREVIMER à destination des chercheurs et de la communauté scientifique. Ces données sont visibles sur le site du projet MARC opéré par l'UMR Laboratoire d'Océanographie Physique et Spatiale : <http://www.umr-lops.fr/Donnees/MARC>

Les modèles MARC se déclinent en plusieurs zones géographiques abordant différentes thématiques. Les chaînes de modélisation sont exécutées quotidiennement sur Datarmor.

### 6.3.2.1 MARC\_L1-MARS2D-SURGE : modèle 2D de courants barotropes, de niveaux et de surcotes

Ce modèle couvre la Mer du Nord et le Golfe de Gascogne (rang 0, résolution 2 km), la façade Manche Gascogne (rang 1, résolution 700 m), et dispose de 5 zooms côtiers (rang 2, résolution 250 m) : Manche Est, Manche Ouest, Finistère, Sud Bretagne et Aquitaine.

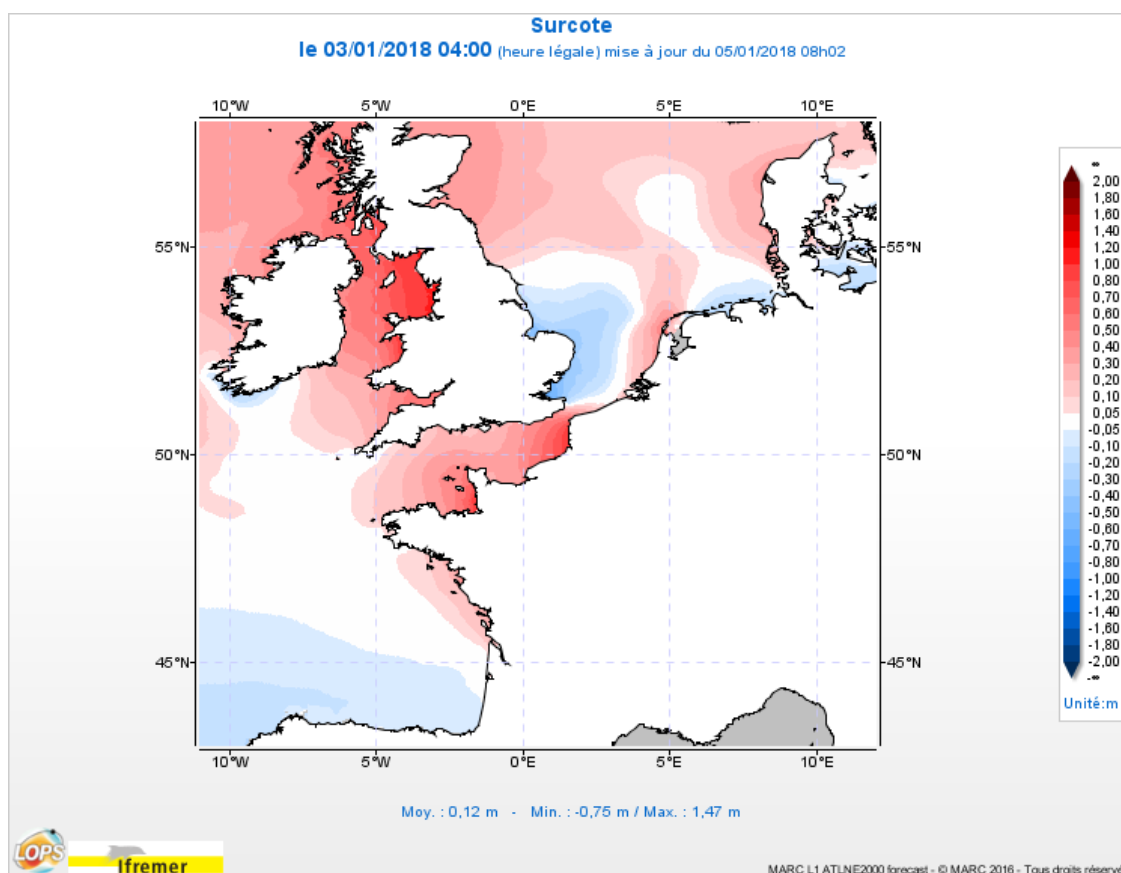


Figure 6-40 : Surcote sur les îles Britannique et la Manche due à la tempête Eleanor du 3 janvier 2018 (extrait du site web [marc.ifremer.fr](http://marc.ifremer.fr))



### 6.3.2.2 MARC\_F1-MARS3D-MANGAE2500 : modèle 3D de courants, température et salinité pour la Manche et le Golfe de Gascogne

Ce modèle de résolution 2,5 km est basé sur la version 9 du code de calcul MARS et utilise les forçages Mercator Ocean aux limites et les niveaux sigma généralisés. La zone couverte est la façade Manche-Atlantique.

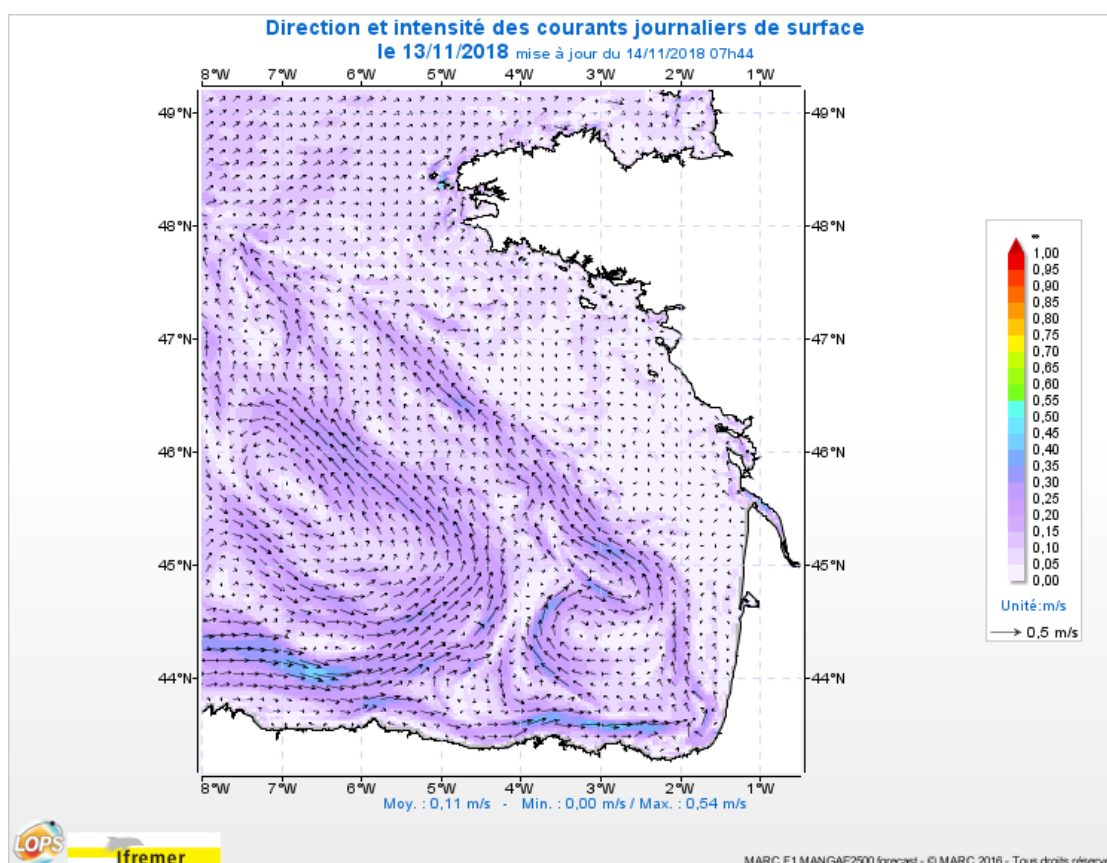


Figure 6-41 : Circulation hivernale dans le Golfe de Gascogne (extrait du site web marc.ifremer.fr)

### 6.3.2.3 MARC\_F1-MARS3D-MANGAE2500-AGRIF : modèle 3D de courants, température et salinité pour la Manche et le Golfe de Gascogne emboîtés « two-way »

Les modèles côtiers ont une résolution de 500m sur toute la façade Manche-Atlantique. Mise en production et alimentation du jeu de données en 2016-2017.

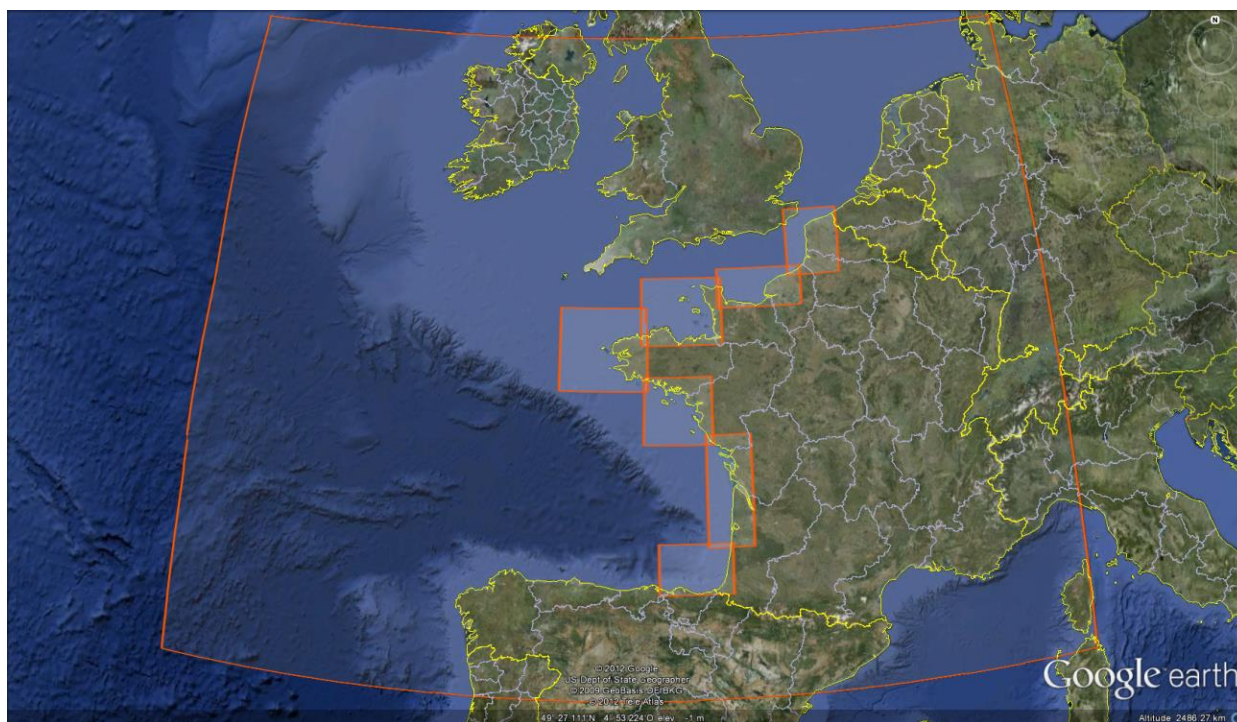


Figure 6-42 : Emprise du modèle Manche-Gascogne à 2,5 km et des ses 7 zooms AGRIF

### 6.3.2.4 MARC\_F2-MARS3D-MENOR1200 : modèle 3D de courants, température et salinité pour la Méditerranée

Ce modèle de résolution 1,2 km cible la façade Méditerranée Nord Occidentale et se base sur la version 10 du code de calcul MARS. Il est forcé aux limites par les données fournies par Copernicus Marine Environment Monitoring Service (CMEMS).

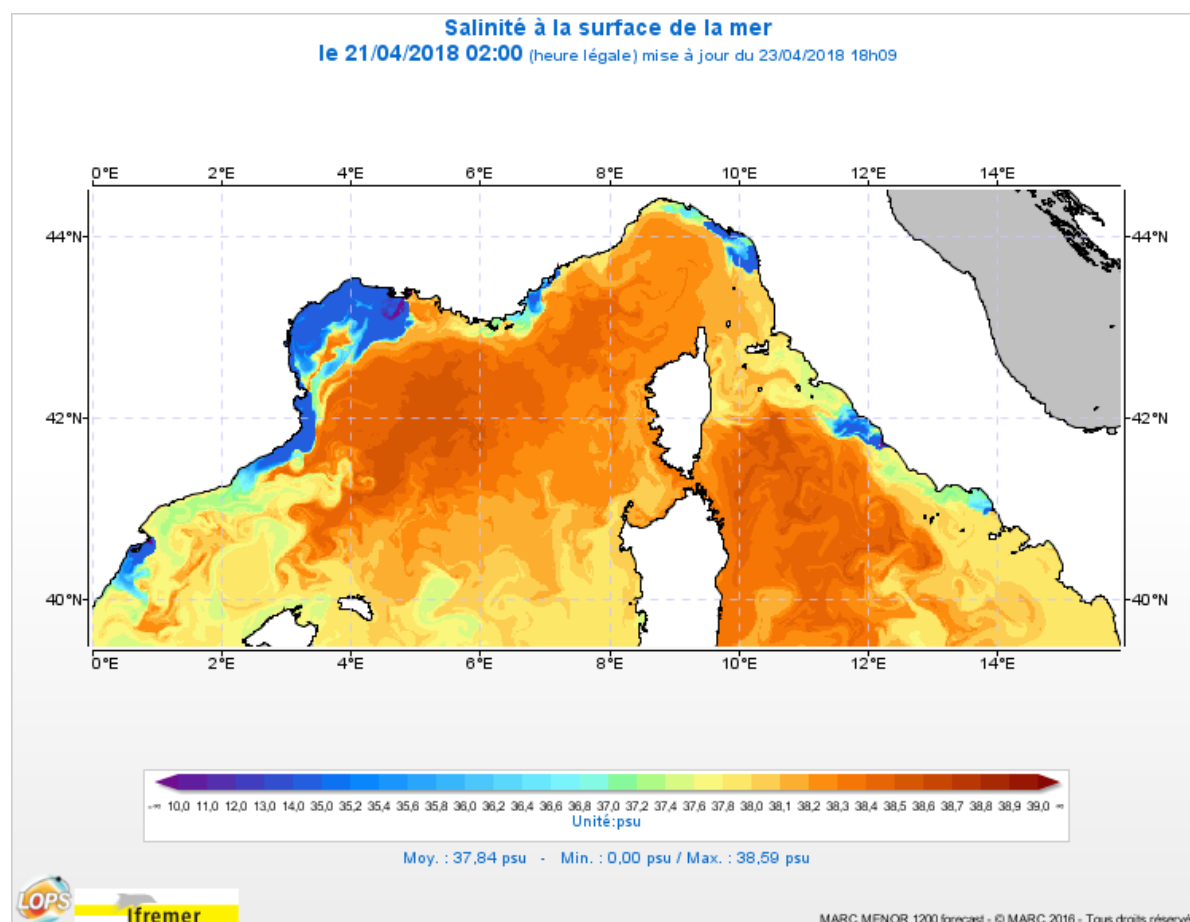


Figure 6-43 : Important dessalure dans le Golfe du Lion suite à un épisode méditerranéen les 11 et 12 avril 2018 (extrait du site web [marc.ifremer.fr](http://marc.ifremer.fr))

### 6.3.2.5 MARC\_B1-ECOMARS3D-MANGA4000 : modèle de production primaire pour la Manche et le golfe de Gascogne

Ce modèle concerne la « production primaire » dans le golfe de Gascogne et la Manche, sa résolution est de 4 km.

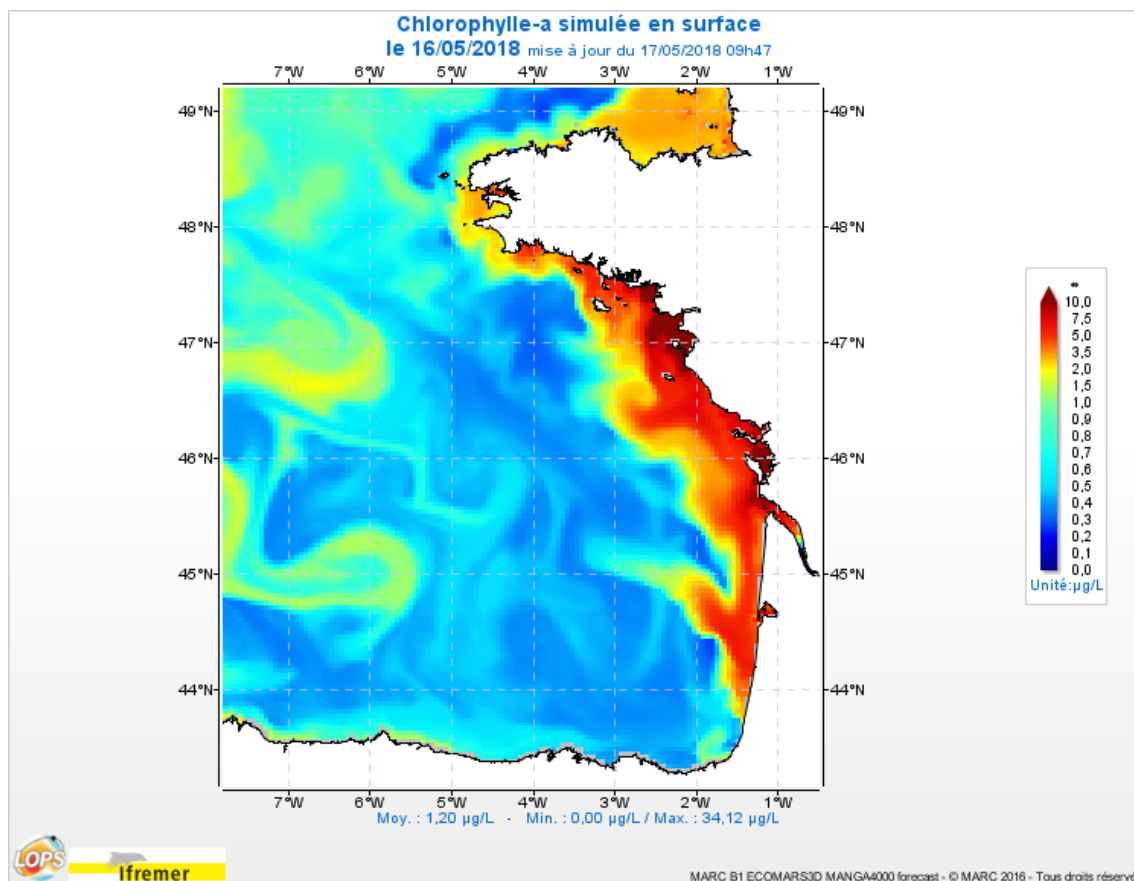


Figure 6-44 : Chlorophylle simulée le 16 mai 2018 (extrait du site web marc.ifremer.fr)

### 6.3.2.6 MARC\_WW3\* : modèles de vagues

Ces modèles sont tous basés sur le code WAVEWATCH III®. Les modèles MARC\_WW3-Globmulti et MARC\_WW3-Medmulti permettent de gérer plusieurs échelles allant du global au régional. Le premier modèle couvre ainsi l'océan global avec une résolution de 30 minutes et les zones ultra-marines (Polynésie, Nouvelle Calédonie et Antilles). Les modèles non structurés emboîtés dans les deux premiers ont des résolutions variables atteignant 200 m près des côtes. Ils couvrent le Golfe de Gascogne, la Manche et la Mer du Nord (Norgas-Ug), la Méditerranée nord occidentale (Menor-Ug), les Antilles (Antilles-Ug) et enfin la Réunion (Reunion-Ug).

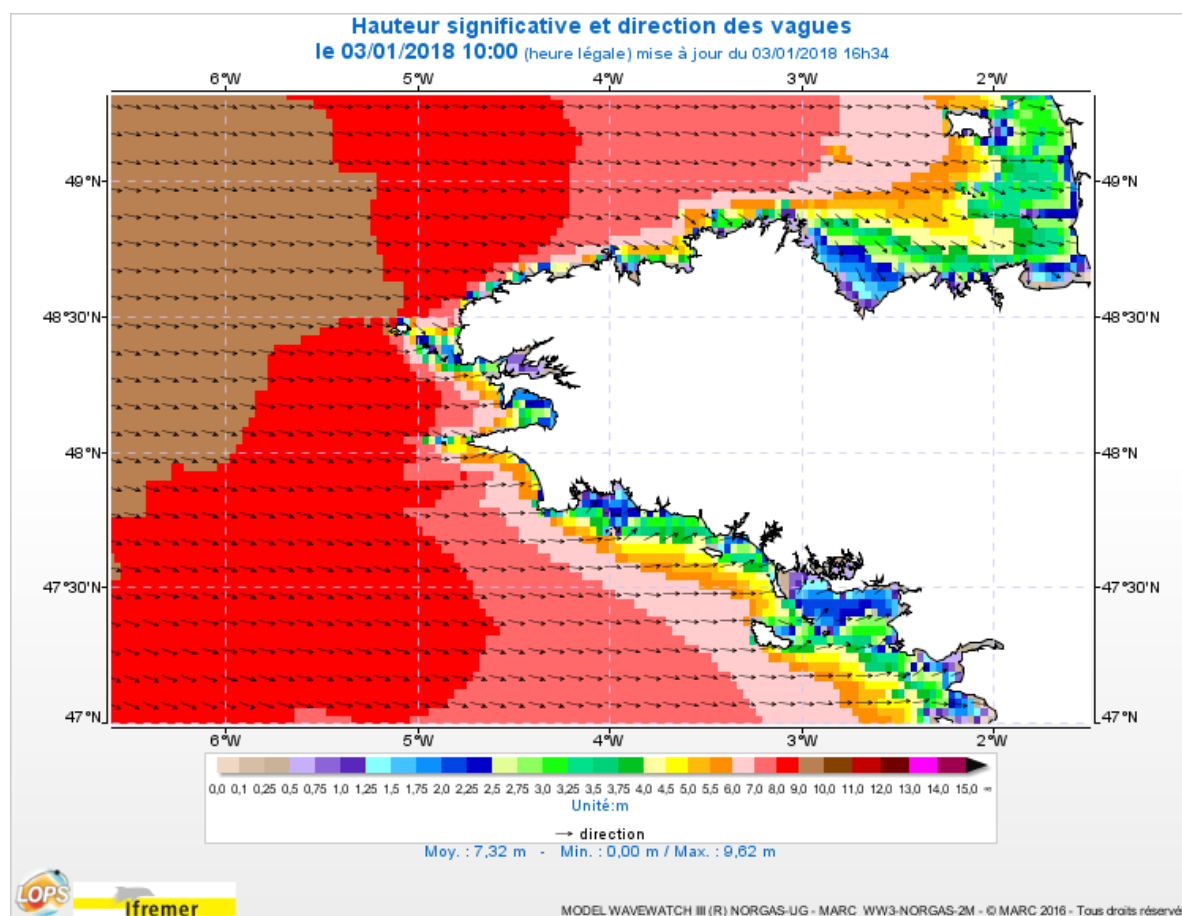


Figure 6-45 : Hauteur significative et direction des vagues le 3 janvier 2018 : tempête Eleanor (extrait du site web [marc.ifremer.fr](http://marc.ifremer.fr))

### 6.3.2.7 MARC\_L2-CORSE400 : modèle 3D de courants, température et salinité pour la Corse

Il s'agit d'un modèle R&D à haute résolution (400m) permettant d'affiner la connaissance des processus hydrodynamiques et d'étudier les transports par les courants.

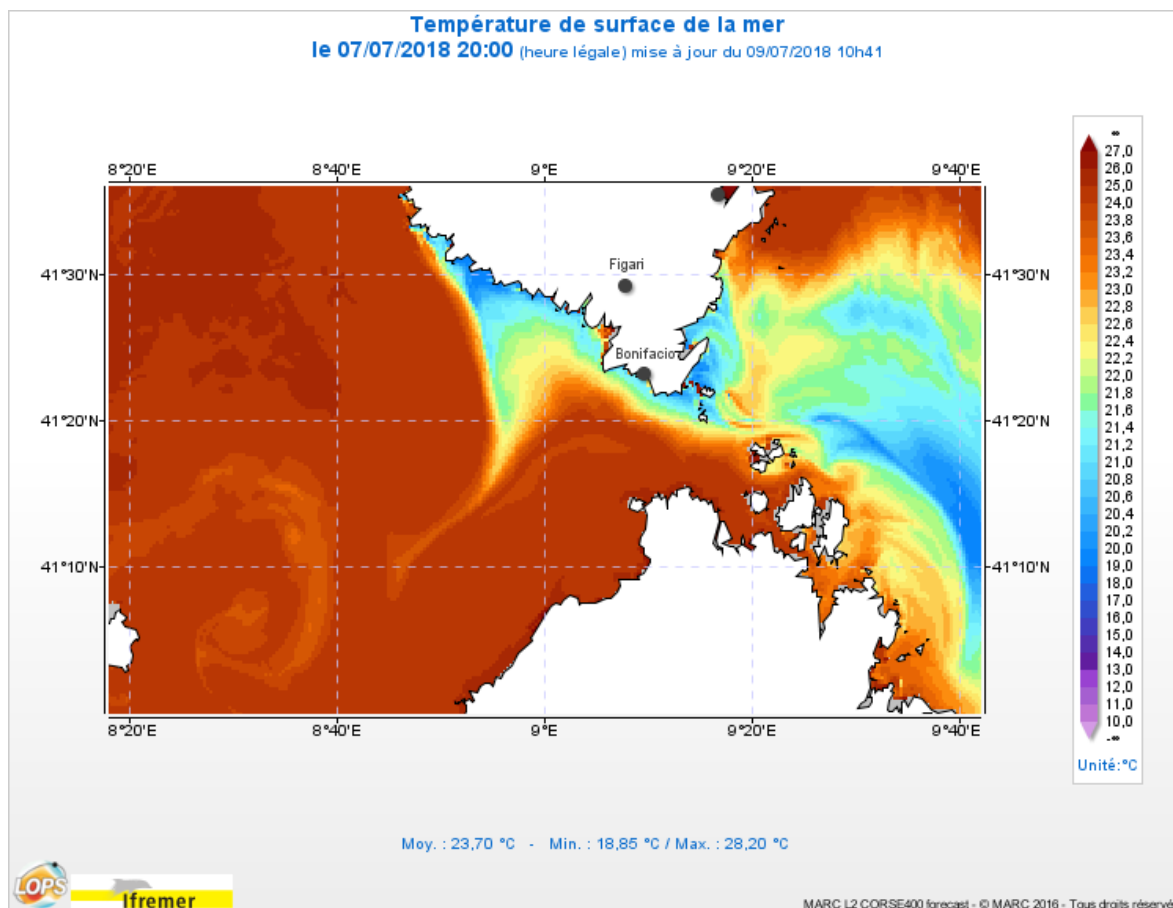


Figure 6-46 : Baisse de la température de surface dans les Bouches de Bonifacio le 7/07/2018 suite à un coup de vent (extrait du site web marc.ifremer.fr)

### 6.3.2.8 MARC\_L1-MARS3D-BREST : modèle 3D de courants, température et salinité pour la Rade de Brest

Il s'agit d'un modèle R&D à haute résolution spatiale (50m) et temporelle (15'), développé pour le suivi environnemental des paramètres physiques (température et salinité), l'estimation du temps d'immersion des huîtres creuses sur l'estran, le suivi larvaire ou de la dispersion de virus, les rejeux d'années sensibles (crises de 2008 et 2009).

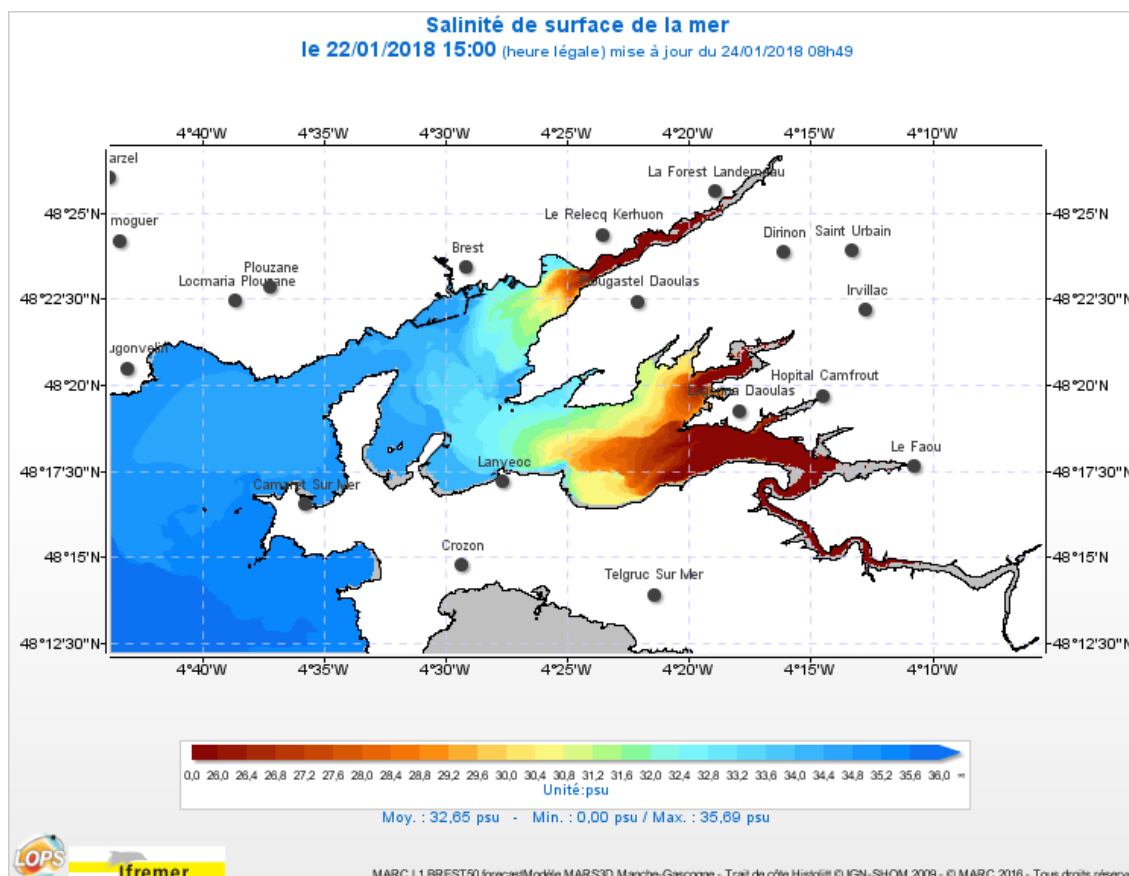


Figure 6-47 : Salinité de surface en rade de Brest le 22/01/2018 (extrait du site web marc.ifremer.fr)

### 6.3.3 Centre opérationnel de données satellite CERSAT

Depuis 2017, le centre CERSAT, précédemment hébergé sur des serveurs de laboratoire, est peu à peu porté sur Datarmor et les infrastructures centrales d'Ifremer. L'opération des flux de données quotidiens est assuré par le centre de données SISMER avec l'appui d'un contrat de sous-traitance.

#### 6.3.3.1 Données

En 2018, les données du CERSAT ont été portées sur Datarmor (dataref, datawork). L'utilisation de ces données peut se faire sur ou en dehors de Datarmor. Des évolutions sont en cours sous Sextant pour permettre une meilleure description de ces jeux de données.

### 6.3.3.2 Portage des outils de processing

Nous avons repris le Maintien en Conditions Opérationnelles (MCO) des outils de processing gérés historiquement par le LOPS :

- **downloader** (python 2) : outil permettant de collecter des données via différents protocoles (ftp, http, openSearch, sftp, ...) et d'effectuer quelques opérations avant de les stocker sur des espaces internes (décompression, création de liens symboliques, ...)
- **chain\_scheduler** (python 2) : outil permettant de lancer des chaînes de traitements (data driven). Les processeurs sont exécutés sur des hôtes distants (vm, Datarmor).
- **cron\_monitor** (python 2) : outil permettant de mettre à jour la crontab avec une liste de commandes à exécuter

Ces trois outils génèrent des logs et des rapports d'exécution. Ces derniers sont moissonnés dans une base MySQL via un traitement nommé **emm\_controller** (python 2). Une **console opérateur** (python 2 + flex + MySQL) permet de consulter les traitements en cours ainsi que les rapports d'exécution (dont notamment les problèmes).

Un gros travail de refonte de ces outils a été réalisé pour simplifier l'installation et l'administration mais aussi pour améliorer la robustesse et la maintenance du code :

- Passage de python 2 à python 3
- Amélioration du code : retrait du code mort, application des bonnes pratiques python, ...
- Simplification de la configuration (ex : 1 seul fichier au lieu de 8 pour le chain\_scheduler)
- Simplification du déploiement :
  - Utilisation du gestionnaire de package « pip »
  - Utilisation d'un serveur Nexus pour déposer les binaires
  - Installation dans des environnements conda
- Ainsi que d'autres évolutions (ex : possibilité de pousser les rapports sous Elasticsearch pour permettre l'utilisation de tableaux de bord via Kibana)
- Supervision des services opérationnels avec systemd, nagios et grafana

En 2018, la console opérateur ainsi que le downloader ont été installés en production dans l'infrastructure RIC. Les autres composants seront installés dans le début de l'année 2019. Les chaînes de traitements CFOSAT pourront tourner intégralement sur Datarmor.

### 6.3.3.3 Portage des outils de diffusion

Les outils de diffusion Naiad et Felyx nécessitent les briques Elasticsearch et RabbitMQ pour fonctionner.

Un cluster Elasticsearch sécurisé a été installé par RIC (achat de matériels dédiés ainsi que l'achat de la solution SearchGuard pour protéger l'accès).

L'installation d'un cluster RabbitMQ se fera en début d'année 2019 via une prestation.



#### 6.3.3.4 Projet Pi-MEP

La mise en place d'une solution JupyterHub sur Datamor permet aux utilisateurs d'exécuter du code en utilisant un environnement de développement Web (sans connexion ssh ni usage de ligne de commande).

Nous avons également déployé l'application Syntool. Cette application permet de consulter des données grillées ou vectorielles via des services au standard géographique WMTS.

Enfin, nous avons installé un environnement Hadoop avec Hbase afin de préparer le portage de l'application Datalaps. En raison du caractère « spécifique » de cette application et des difficultés à trouver des informations, nous avons commencé à étudier des alternatives à partir d'outils développés et maintenus par la NASA.

## 7 Configuration technique de DATARMOR

La société SGI, reprise par HPE en 2017, a fourni une solution adaptée et performante en réponse aux besoins des utilisateurs scientifiques de l'Ifremer et de ses partenaires, pour le calcul et pour le stockage. Une évolution de la solution a été réalisée en 2018.

Les services proposés sur Datarmor s'appuient sur différentes composantes.

### 7.1 Architecture générale

Datarmor propose des solutions spécifiques de calcul associées à des solutions de stockage. Un réseau très performant relie les différentes composantes, tant au sein de la machine qu'avec le réseau de l'Ifremer.

Afin d'obtenir le meilleur rendement PUE, l'organisation des baies informatiques (et donc l'architecture) sépare bien les éléments de calcul/traitement (qui disposent d'un refroidissement à eau spécifique) des éléments dédiés au stockage (qui sont refroidis par un système hybride air/eau).

Les éléments dédiés au traitement/calcul sont :

- l'ICE XA constitué de 396 nœuds de calcul pour un total de 11088 coeurs, principalement dédié au calcul parallélisé,
- l'UV3000 fournissant 5 To de mémoire non distribuée, donc, principalement dédiée aux travaux demandant une grande quantité de mémoire et ne pouvant utiliser la mémoire distribuée de l'ICE XA,
- une composante GPU sur deux serveurs avec un total de quatre GPU Nvidia Quadro K6000,
- une composante de dix serveurs pour fournir des services type web, base de données, ...

Les éléments dédiés au stockage sont :

- un espace dédié aux données de travail (datawork) géré par le système de fichier GPFS totalisant 5,7 Pto utiles,
- un espace dédié aux données de référence (dataref) géré par le système de fichier lustre totalisant 1,395 Pto,
- un espace dédié aux données des comptes utilisateurs (datahome) géré par système de fichier xfs totalisant 37 Tto,
- un espace dédié aux données temporaires (scratch) géré par le système de fichier lustre totalisant 522 Tto,
- un espace dédié aux services (web, sgbid, ...) totalisant 135 Tto.

Ces éléments sont reliés entre eux par un réseau infiniband à 56 Gbits, en double plan. Les accès depuis le réseau ifremer ou externe se font par un réseau à 10 Gbits permettant des accès en NFS (unix) et CIFS (windows).

Le matériel informatique de Datarmor prend place dans une infrastructure comprenant

- la salle machine sécurisée (accès, détection et extinction incendie), un local onduleur, une plateforme extérieure ainsi qu'un transformateur (TGBT),
- un onduleur de 320 KVA et batteries associées,
- un groupe électrogène de 550 kVA permettant une reprise en pleine charge avant une minute,
- une climatisation des serveurs par trois inrows (salle machine) avec condenseurs (plateforme extérieure),
- une dissipation de la chaleur de l'ICE XA par dry cooler (plateforme extérieure) redondée/secondée par l'eau glacée de site au travers d'un échangeur.

## 7.2 Les clusters d'exécution

L'exécution des travaux des utilisateurs est orchestrée par le gestionnaire de tâches PBS pro. C'est ce gestionnaire qui redirige les « jobs » sur la partie ICE XA (traitements MPI), l'UV3000 ou les serveurs ayant des GPU (traitements non MPI). Ce gestionnaire est directement accessible à partir des quatre nœuds de login proposés sur Datarmor.

**L'ICE XA** dispose de 396 nœuds de calcul contenus dans deux racks. Chacun de ces nœuds dispose de deux processeurs (E5 2680 v4 à 2.4 GHz), chaque processeur ayant 14 coeurs. Ainsi,



Figure 7-1: lame contenant 4 nœuds de calcul, refroidissement liquide des CPU

chaque nœud propose 28 coeurs de calcul, le double en s'appuyant sur la technologie intel de l'hyperthreading. La puissance de calcul crête de cette solution est de 426 Tflops. Chaque nœud de calcul comprend 128 Go de mémoire DDR4 2400 MHz, totalisant ainsi 50 To de mémoire distribuée pour l'ICE. Chaque nœud est connecté à la « fabric infiniband » par deux liens de 56 Gb/s chacun. La configuration en place utilise un lien pour les échanges de données (accès disque) et un lien pour les échanges entre les nœuds (communication MPI). L'ICE XA est associé au

refroidissement selon un agencement très dense : 4 nœuds de calcul par lame. La dissipation des calories se fait directement au niveau de chaque CPU par liquide, le refroidissement global (mémoire, carte mère, ...) est lui fait par air, l'ensemble de l'ICE étant dans un compartiment fermé. La chaleur produite par l'ICE est gérée par un CDU qui assure l'échange des calories avec l'installation extérieure ainsi que la circulation et la pression du fluide dans l'ICE XA.

**L'UV3000** offre une capacité importante de mémoire non distribuée : 5 To (DDR4 2133MHz). Cette machine est constituée de 10 lames assemblées en une solution vue comme unique, destinée à des traitements non MPI. Chaque lame comprend deux processeurs Intel Xeon E5-4650v3 à 2.1 Ghz, chacun ayant 12 coeurs de calcul. Cette solution correspond, du point de vue de l'utilisateur, à une machine de 240 coeurs et 5 To de mémoire pour une puissance de calcul de 8 TFlops. Les échanges avec les autres composants de Datarmor (stockage disque) se font par le réseau infiniband à 56 Gb/s. La chaleur dégagée par l'UV3000 est évacuée par les inrows de la salle machine qui proposent en face avant de l'air frais.



*Figure7-2 : UV3000, 10 lames*



*Figure 7-3 : Nvidia Quadro K6000*

Les deux **serveurs de visualisation** ont chacun deux processeurs Intel Xeon E5-2630v4 2,2GHz, chaque processeur ayant dix coeurs. Ils disposent de 128 Go de mémoire DDR4 2400MHz. Ils sont connectés (stockage disque) avec un lien infiniband à 56 Gb/s. Ils embarquent chacun deux cartes graphiques Nvidia Quadro K6000, chacune proposant 448 coeurs cuda. La chaleur dégagée par ces serveurs est évacuée par les inrows de la salle machine qui proposent en face avant de l'air frais.



*Figure 7-4 : serveurs de service*

Les dix **serveurs de services** (web, données, ...) sont des bi-processeurs Intel Xeon E5-2620v3 cadencé à 2,4GHz, chaque processeur ayant 6 coeurs de calcul. Ils contiennent 64 Go de mémoire DDR4 à 2133MHz. Ils sont tous connectés au réseau Ifremer par des ports 10 Gb/s, cinq sont également connectés à l'ensemble Datarmor (stockage) par un lien infiniband à 56 Gb/s. Ils sont directement connectés au stockage dédié par fibre channel. La chaleur dégagée par ces serveurs est évacuée par les inrows de la salle machine qui proposent en face avant de l'air frais.

## 7.3 Les clusters de stockage

Plusieurs espaces sont proposés sur la plateforme Datarmor avec des rôles spécifiques :

- **datawork** dédié aux données de travail est un espace de 5,7 Pio. Il a fait l'objet d'une extension en 2018, disques et enclosures. Il s'appuie sur une appliance DDN pour le système de fichier GPFS, appliance composée de deux contrôleurs redondants reliés à dix tiroirs de disques. L'ensemble contient 1060 disques SAS de 8 To à 7200 tr/mn ainsi que 34 disques SDD de 800 Gb. Cette solution de stockage permet des accès à un débit effectif de 30 Go/s en lecture/écriture parallèle. Cet espace ne dispose pas d'une sauvegarde sur bande mais peut, pour des contextes spécifiques, mettre en œuvre des instantanés (snapshot).



Figure 7-5 : tiroir de disque du GPFS

- **dataref** est un espace d'accueil des données de référence, 1,395 Pio sont proposés. Cet espace est supporté par le système de fichier lustre avec un serveur d'administration, deux serveurs de meta-donnée et deux serveurs de donnée. Ces serveurs sont reliés à Datarmor par des liens infiniband à 56 Gb/s. Le stockage des méta-données se fait sur 12 disques SAS de 1.2 To à 10000 tr/min dans une baie de stockage IS5100, le stockage des données se fait sur 240 disques de 8 To SAS à 7200 tr/min avec gestion par un contrôleur redondé. Cette solution de stockage permet des accès à un débit effectif de 25 Go/s en lecture/écriture parallèle.



Figure 7-6 : contrôleurs et baies disques des espaces lustre

- **scratch** est un espace d'accueil temporaire des données produites, 522 To sont proposés. Cet espace est supporté par le système de fichier lustre avec un serveur d'administration, deux serveurs de meta-données et quatre serveurs de données. Ces serveurs sont reliés à Datarmor par des liens infiniband à 56 Gb/s. Le stockage des méta-données se fait sur 12 disques SAS de 1.2 To à 10000 tr/min dans une baie de stockage IS5100, le stockage des données se fait sur 180 disques de 4 To SAS à 7200 tr/min avec gestion par un contrôleur redondé. Cette solution de stockage permet des accès à un débit effectif de 25 Go/s en lecture/écriture parallèle. Les données présentes sur cet espace sont effacées au bout de 15 jours.



Figure 7-7: serveurs et stockage des home directories

- **datahome** est l'espace des répertoires des utilisateurs, les « home directories ». Cet espace de 37 Tio est supporté par deux serveurs redondants reliés à une baie de stockage (IS5100) disposant de 24 disques de 2 To à 7200 tr/mn. Ces serveurs sont reliés à Datarmor par des liens infiniband à 56 Gb/s, au réseau Ifremer par des liens ethernet à 10 Gb/s. Cet espace est sauvegardé chaque nuit avec une durée de rétention des sauvegardes de deux mois.

- **dataweb** est un stockage de 135 Tio associé aux serveurs de service tels que le web. Il est contenu dans une baie IS5600i avec 50 disques de 4 To à 7200 tr/mn ainsi que 2 disques SSD de 400 Go avec une licence « flash cache ».
- L'accès aux données de datawork, dataref et scratch, depuis le réseau Ifremer, se fait au travers de **passerelles** redondées proposant NFS (accès depuis des machines unix) et CIFS (accès depuis des machines windows). Les deux serveurs dédiés au stockage lustre (scratch et dataref) sont des bi-processeurs Intel Xeon E5-2620v4 cadencés à 2,1GHz ayant 8 coeurs par processeur et 64 Go de mémoire DDR4 à 2400MHz. Les deux serveurs en cluster dédiés au stockage GPFS (datawork) sont des HPE ProLiant DL360 Gen10 ayant chacun un processeur Intel Xeon-Gold 6140, 18 cores 2.3GHz et 192Go mémoire DDR4 RDIMM 2666MHz. Ils sont reliés à Datarmor lien infiniband à 56 Gb/s et reliés au réseau Ifremer lien ethernet 10 Gb/s.



Figure 7-8 : stockage associé au web, un tiroir sorti

Les espaces de stockage dégagent une chaleur importante. Les calories sont évacuées par les trois inrows de la salle machine qui soufflent un air refroidi en face avant des racks.

Les tailles des espaces sont exprimées en Tio ou Pio, le facteur multiplicateur est donc 1024 et non 1000. Cela correspond aux valeurs par défaut fournies par la commande df.

## 7.4 Les interfaces et les réseaux



Figure 7-9- : switches infiniband et cablage dans l'ICE XA

Les constituants de Datarmor sont reliés par différents réseaux. Celui permettant les bonnes performances de la machine est l'infiniband FDR à 56 Gb/s. Les liens avec le réseau Ifremer sont en ethernet à 10 Gb/s. Un réseau s'appuyant sur des brocades et des liens fibre channel relie l'espace dataweb aux serveurs de service et permet la sauvegarde journalière des répertoires des utilisateurs. Enfin, un réseau spécifique d'administration permet le suivi et la gestion des composantes de la machine.

Le **réseau infiniband** s'appuie sur les switches infiniband compris dans l'ICE XA, 4 switches dans chaque IRU des racks. Ces switches permettent de relier les nœuds de calcul entre eux mais également de connecter les composantes hors de l'ICE XA : serveurs dédiés au stockage, composantes de traitement non MPI, serveurs pour les services web, nœuds de login, passerelles. Ce réseau est en double plan : un plan pour les échanges dédiés aux données (stockage) et un plan pour les échanges entre machines dans le cas de traitement MPI. La topologie mise en œuvre est « enhanced hypercube », solution optimisant les échanges entre les matériels dans le cadre d'un réseau double plan avec une meilleure tolérance aux pannes.

Le **réseau 10 Gb/s** s'articule autour de deux commutateurs réseau Extreme Networks X670-G2 48 ports à 10 Gb/s, chacun ayant deux liens avec le cœur de réseau Ifremer à 40 Gb/s. Ce réseau propose un accès performant à l'architecture Datarmor depuis le réseau Ifremer. C'est le cas par exemple des serveurs de répertoires utilisateur (datahome) ainsi que des passerelles et des serveurs de service.



Figure 7-10

Certains échanges s'opèrent en s'appuyant sur un **réseau fibre channel** connectés à des brocades, connexion au stockage dataweb et connexion à l'espace de stockage des répertoires des utilisateurs. Le matériel en œuvre est de deux brocades switches 300 avec 24 ports 8 Gb/s



Figure 7-11

Le suivi et la gestion des matériels de l'architecture s'appuient sur un **réseau d'administration** avec deux switches Extreme Network Switch X440 48 ports 1Gb/s.

## 7.5 La logistique

Plusieurs contraintes sont à prendre en compte pour la mise en place d'une architecture matérielle comme Datarmor. Bien sûr, une fourniture électrique continue et fiable pour assurer le bon fonctionnement de matériels. Conséquence de cette consommation, la chaleur générée doit être évacuée, au moindre coût. La sécurisation de l'ensemble du périmètre doit également être en place : contrôle entrées et sorties, détection incendie, extinction incendie.

Il y a plusieurs sources de chaleur dans l'installation Datarmor et ainsi que plusieurs solutions d'**évacuation de la chaleur**.



Figure 7-13 : dry cooler de l'ICE XA

L'ICE XA avec 396 nœuds de calcul est la principale source. Sa conception fait qu'il est thermiquement isolé du reste des installations et de la salle. La reprise des calories se fait directement par eau au niveau des processeurs ou par air dans l'enceinte close de l'ICE XA. Cet échange se fait sur une plage de température haute : 30 / 35 °C. Le CDU assure l'échange calorifique avec le dry cooler sur la plateforme extérieure. Celui-ci, sans action d'un compresseur, évacue la chaleur dans l'air extérieur si la température de l'air est inférieure à 25 °C. Au-delà de cette valeur ou lors d'un arrêt du dry cooler, l'eau glacée du site Ifremer est utilisée en modulant le débit selon le besoin par une vanne trois voies motorisée.



Figure 7-12 : CDU de l'ICE XA

Les autres éléments de Datarmor (stockage, réseau, serveurs, UV3000, ...) génèrent une chaleur en arrière d'une rangée de racks, air chaud repris par les inrows qui soufflent une lame d'air frais en avant des racks. Ces trois inrows, nombre assurant une redondance, utilisent un compresseur et un fluide calo-porteur pour évacuer les calories vers trois condenseurs sur la plateforme extérieure de Datarmor.

Enfin, l'onduleur installé dans une salle technique dédiée, génère également de la chaleur. Une climatisation redondante est en place pour garantir la constance des conditions d'exploitation.



Figure 7-14 : condenseurs associés aux inrows

Le PUE ou Indicateur d'efficacité énergétique constatée pour la plateforme Datarmor est de 1,27, depuis sa mise en exploitation.



**L'apport électrique** est assuré par plusieurs sources : EDF en nominal via la boucle 20 000 V du site. Un TGBT dédié à Datarmor a été mis en place.

La qualité de la fourniture électrique au matériel informatique est assurée par un onduleur de 320 KVA. Cet onduleur, associé à un banc de batteries, permet également le fonctionnement en pleine charge de l'ensemble Datarmor pendant au moins cinq minutes.

Un groupe électrogène de 550 KVA dédié à Datarmor, situé sur la plateforme extérieure, permet la reprise en pleine charge de Datarmor en moins d'une minute (trente secondes lors des tests). Les génératrices du site de Brest peuvent également, par la suite, reprendre la totalité du site dont Datarmor.



*Figure 7-15 : groupe électrogène de Datarmor*

La **sécurisation de l'installation** contre l'incendie est assurée par un ensemble de détecteurs installés de manière dense dans la salle datarmor et dans la salle onduleur. Ces détecteurs sont reliés à une centrale incendie et un ensemble d'éléments d'avertissement. Un système d'extinction par gaz inerte est également installé et couplé à la centrale incendie, dans la salle datarmor et dans la salle de l'onduleur. Des matériels spécifiques sont mis en œuvre en salle datarmor pour protéger les éléments de stockage en cas de déclenchement de l'extinction. Un détecteur de fluide est également en place sous l'installation informatique pour alerter d'une fuite de liquide en salle.

L'accès à la salle datarmor est couplé au système de contrôle des accès mis en place sur les sites de Ifremer. Les outils déjà en place permettent donc de gérer les droits sur les cartes utilisées au quotidien par les agents Ifremer.



*Figure 7-16 : extinction incendie*

## 7.6 Les liaisons réseau inter-organisme

Les liens entre l'Ensta Bretagne, le Shom Brest, l'IUEM et l'UBO sont établis en point à point entre établissements sur la base de fibres noires louées à Brest Métropole, afin de limiter les règles de sécurité à intégrer dans les pare-feux. Le débit de ces liens est de 1 gbit/s, donc équivalent à celui d'un réseau local. Le but du réseau INFUSER est de rendre transparents à l'utilisateur les transferts de données volumineuses.

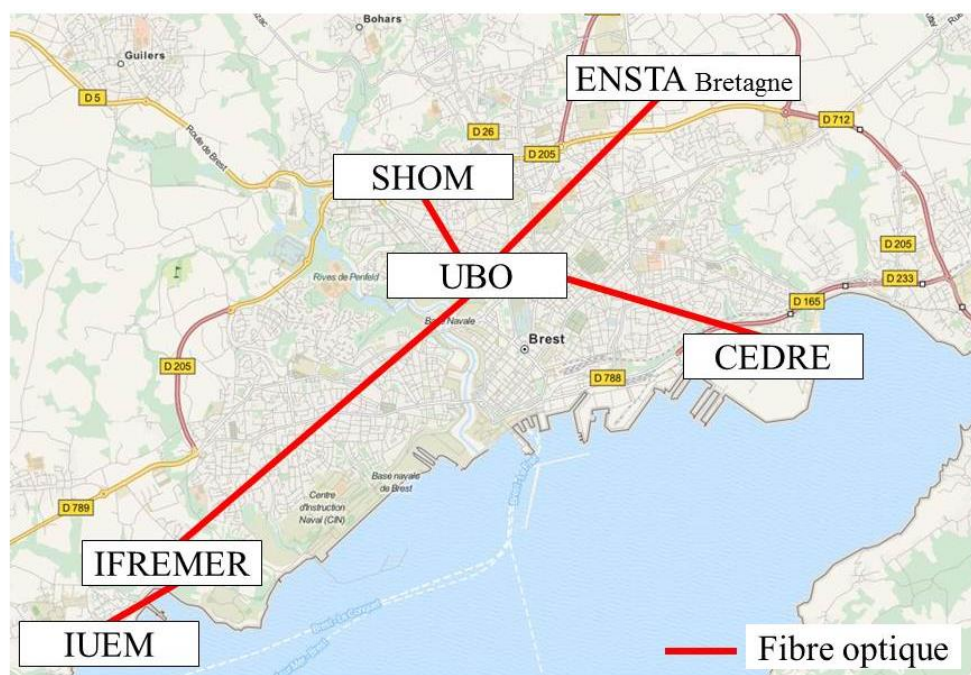


Figure 7-17 : Réseau métropolitain de Brest (réseau INFUSER)

## 7.7 Les principaux logiciels

La plateforme Datarmor propose des ensembles de logiciels nécessaires au large éventail des applications utilisées dans le cadre des sciences marines.

Les utilisateurs peuvent installer leurs ensembles applicatifs dans les espaces disques spécifiques attachés à leur compte sur le calculateur. L'équipe RIC installe, met à disposition et effectue le suivi de logiciels nécessaires à des communautés d'utilisateurs.

Les logiciels de base fournis avec la plateforme matérielle comprennent

- le gestionnaire de tâches PBS pro de la société Altair,
- l'environnement de compilation, débogage et optimisation Intel Parallel Studio XE Cluster Edition dont les performances sur processeurs Intel sont largement reconnues, cas de la librairie mathématique MKL,
- Alinea DDT, débogueur adapté au cas du traitement HPC
- les outils fournis par le constructeur SGI Performance Suite et SGI Management Suite permettant à la fois une utilisation performante de la plateforme mise en œuvre et l'administration d'un ensemble matériel et logiciel complexe.

A ces logiciels, l'équipe RIC propose des logiciels dont le besoin a été identifié et confirmé au cours des années :

- bien sûr, les compilateurs comme le Fortran de PGI, le compilateur gcc pour C, C++ et fortran de Gnu, le langage python utilisé dans de très nombreux développements, le compilateur GO de google qui vise performance et parallélisations,
- des outils pour identifier/traquer les problèmes comme valgrind : recherche d'erreurs de gestion mémoire de codes,
- des applications scientifiques comme R et Octave ainsi que les bibliothèques associées à python comme numpy,
- des applications thématiques comme le sont les logiciels dédiés à la bio-informatique,
- les outils associés au format netCDF comprenant les bibliothèques utilisées lors de la compilation des codes, les outils de manipulation des fichiers (cdo et nco), les outils de visualisation comme nview.

Ces logiciels sont mis en commun mais d'autres sont également installés et proposés en réponse à des besoins parfois plus spécifiques, dans un premier temps au moins. Ces nombreux outils sont sous forme de briques que chaque utilisateur peut lister et, selon ses besoins, mettre en œuvre dans son environnement via des modules unitaires à charger.

## 8 Annexes

### 8.1 Liste des membres du Comité de Pilotage Datarmor

#### Membres titulaires :

M. Pierre GARREAU - IFREMER ODE/LOPS/OC - PRESIDENT DU COMITE

M. Erwan JOUANNET – SHOM

M. Jean-Marc DANIEL - IFREMER REM

Mme Sylvie LE GENTIL - IFREMER ODE/LOPS

Mme Stéphanie MAHEVAS – IFREMER RBE/EMH

M. Jean-François PIOLLE – IFREMER ODE/LOPS

M. Philippe CUGIER - IFREMER ODE/DYNECO

#### Membres invités :

M. François DENISET - ECOLE NAVALE

Mme Anne-Laure TIBERI - CEREMA

M. Nicolas GUILLOU – CEREMA

M. Bruno FREDERIC – ENSTA BRETAGNE

M. Olivier AUMONT – IRD

Mme Laurence LEBOURG– UBO/IUEM

M. Philippe MONBET – POLE MER BRETAGNE ATLANTIQUE

M. Gilbert MAUDIRE - IFREMER IRSI

Mme Anne THOMAS-DOARE - IFREMER IRSI/RIC

M. Denis CROIZE-FILLON - IFREMER IRSI/RIC

M. Dominique LE BRUN - IFREMER IRSI/RIC

M. Patrick DURAND - IFREMER IRSI/RIC/BIO-INFORMATIQUE

M. Pierre COTTY - IFREMER IRSI - SECRETAIRE DU COMITE

## 8.2 Glossaire

ADCP : Acoustic Doppler Current Profiler (instrument de mesure de courant)

ARCHIMER : Archive institutionnelle de l'Ifremer

BLAST : Basic Local Alignment Search Tool

CATDS : Centre Aval de Traitement des Données du satellite SMOS

CEREMA : Centre d'études et d'expertise sur les risques, l'environnement, la mobilité et l'aménagement

CERSAT : Centre de données satellite Ifremer

CFOSAT : Satellite franco-chinois

CIFS : Common Internet File System

CLUSTER : (informatique) Grappe/groupement de serveurs techniquement associés pour ne former une entité unique

CNES : Centre National d'Etudes Spatiales

CNRS : Centre National de la Recherche Scientifique

COI : Commission Océanographique Internationale (émanation de l'UNESCO/ONU)

CONDA : outil/logiciel de conteneur virtualisé

COPERNICUS : Copernicus est le programme européen pour l'observation de la terre coordonné par la Commission Européenne

CPER : Contrat de Plan Etat-Région

CPU : central processing unit

DATARMOR : Infrastructure de calcul/stockage/traitement de données marines gérée par l'Ifremer

DATA TERRA : Infrastructure de Recherche (IR) labellisée par le MESRI (anciennement Pôles de données du système Terre)

DCE : Directive Cadre sur l'Eau

DCSMM : Directive Cadre de Stratégie sur le Milieu Marin

DHFF : Directive Habitat-Faune-Flore

DIAS : Cloud UE de services d'accès aux données et à l'information

DOCKER : outil/logiciel de conteneur virtualisé

DOI : Digital Object Identifier

DYNECO : Unité de l'Ifremer – DYNamiques des Ecosystèmes COtiers

ELASTICSEARCH : moteur de recherche et d'analyse

EMODNET : European Marine Observation and Data Network (réseau intégré d'organisations pour des services de données marines)

ENSTA : Ecole Nationale Supérieure de Techniques Avancées

EOSC : European Open Science Cloud

FAIR Data : Findable, Accessible, Interoperable, Reusable Data

FDR : Fourteen Data Rate (infiniband)

GALAXY : Portail Web intégrant des logiciels de bio-informatique

GLIDER : Sorte de planeur sous-marin autonome et télécommandé, capable de faire des mesures à diverses profondeurs

GPFS : General Parallel File System

GPU : graphics processing unit

HPC : High Performance Computing (=calcul intensif)

HPE : Hewlett Packard Enterprise

ICE : Integrated Cluster Environnement

INSPIRE : Directive européenne visant à créer une infrastructure intégrée de données géographiques en support de la politique environnementale européenne

IODE : Intergovernmental Oceanographic Data and Information Exchange (entité de la COI/UNESCO faisant la promotion de la mise en commun des données de la recherche océanographique)

IR : Infrastructure de Recherche (label donné par le MESRI)

IRD : Institut de Recherche pour le Développement

IRU : Individual Rack Unit

ISO 19115 : norme de référence pour l'information géographique dans le domaine des métadonnées.

IUEM : Institut Universitaire Européen de la Mer – Composante de l'UBO

JUPYTERHUB : ensemble logiciel permettant l'accès à un environnement de développement avec une interface Web

KIBANA : permet de visualiser les données Elasticsearch et de naviguer dans la Suite Elastic

LOPS : Laboratoire d'Océanographie Physique et Spatiale (UMR)

LUSTRE : système de fichiers parallèle d'origine opensource

MARBEC : MARine Biodiversity, Exploitation and Conservation, unité de recherche qui regroupe des personnels de 4 organismes : l'IRD, l'Ifremer, l'Université de Montpellier et le CNRS.

MARC : Modélisation et Analyse pour le Recherche Côtière

MESRI : Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation

Météo-France : Agence Nationale de Météorologie

MPI : Message Passing Interface

NAIAD : logiciel Ifremer permettant l'accès et la visualisation en interface Web sur des données images

NetCDF : Network Common Data Form

NEMO : Logiciel Ifremer de formatage de données créé dans Seadatanet

NFS : Network File System

NODC : National Oceanographic Data Center

NoSQL : Not Only SQL (organisation de données alternative aux bases de données relationnelles)

OCTAVE : logiciel GNU de calcul numérique

ODATIS : Pôle océan, partie intégrante de l'IR « pôles de données du système Terre » - Data Terra

ODE : Département de l'Ifremer - Océanographie et Dynamique des Ecosystèmes

OPENSOURCE : Opensource est une approche non commerciale pour le design, le développement et la diffusion de logiciel

PALMIRA : Projet Ifremer (Paléoécologie d'ALexandrium MInutum en RAde de Brest)

PUE : Power Usage Effectiveness

PYTHON : langage de programmation (script)

R : logiciel opensource de calcul statistique et visualisation, associés à de nombreux paquets numériques

RAM : "Random Access Memory" : mémoire vive d'un ordinateur

RIC : Service de l'Ifremer - Ressources Informatiques et Communications

SEADATACLOUD : Projet européen suite de Seadatanet2, coordonné par Ifremer

SEADATANET : Projet européen coordonné par Ifremer mettant en réseau les centres européens de données marines

SEXTANT : Projet Ifremer (et par extension outils afférents) basé sur les techniques de géoréférencement

SIG : Silicon Graphic International

SHOM : Service Hydrographique et Océanographique de la Marine (Organisme dépendant du Ministère de la Défense)

SMP : Shared Memory Processing : technologie de serveurs à mémoire partagée.

SGBD : Système de Gestion de Base de Données

SISMER : Service de l'Ifremer - Systèmes d'Informations Scientifiques pour la MER

TGBT : Tableau Général Basse Tension

UMR : Unité Mixte de Recherche

*Impression : Service IRSI-RIC  
IFREMER – Centre de Bretagne  
CS 10070 – 29280 Plouzané  
Tél. : 02 98 22 43 53*