



# Les Grands Dauphins du golfe normand-breton : facteurs environnementaux déterminant leur distribution et modélisation de leur niche écologique

ASSOCIATION AL LARK



Du 26/02/2020 au 26/08/2020

Sous la direction de : ESTELLE PETIAU

*Le présent rapport constitue un exercice pédagogique qui ne peut en aucun cas engager la responsabilité de l'entreprise ou du laboratoire d'accueil*

Oihana Olhasque

Master 2 DynEA – Année universitaire 2019/2020

## Table des matières

Liste des sigles / Liste des illustrations .....	
Remerciements .....	
Avant-propos .....	
L'association.....	
Déroulement des 6 mois de stage .....	
I. Introduction.....	1
II. Matériel et méthodes.....	4
1. Site d'étude .....	4
2. Le Grand Dauphin Tursiops truncatus .....	6
3. Les données collectées .....	7
3.1. Généralités .....	7
3.2. Données collectées par suivis scientifiques .....	7
3.3. Données supplémentaires.....	8
4. Analyses préliminaires sur la période 2016-2019 .....	10
5. Analyse de la distribution spatiale des Grands Dauphins .....	11
5.1. Préparation des données .....	12
5.2. Choix des types de modèles .....	13
III. Résultats .....	17
1. Reformatage des données et incrémentation de la base de données existante.....	17
2. Résultats des analyses préliminaires.....	18
3. Modélisation.....	20
Résultats de l'ENFA .....	21
Résultat des GLM /GAM .....	24
IV. Discussion .....	27
Bibliographie.....	31
Table des annexes.....	35
Annexes.....	
Résumé .....	
Abstract.....	
Resumen .....	

## Remerciements

Par ce présent rapport, je tiens à remercier le directeur de l'Association AL LARK, Gaël GAUTIER, de m'avoir accueillie durant ces six mois de stage au sein de son association, et d'avoir su se montrer disponible, patient, et souriant en toute situation.

Je remercie Estelle PETIAU, ma tutrice de stage, de m'avoir fait confiance, de m'avoir guidée et de m'avoir permis de réaliser ce stage dans les meilleures conditions possibles. Je tiens à lui exprimer mes sincères remerciements pour son écoute, ses conseils et pour avoir su partager sa passion avec moi.

Je tiens aussi à remercier Morgane PERRI, pour ses conseils, son professionnalisme et l'attention dont elle a fait preuve à mon égard durant ce stage.

Je remercie vivement Jérémy DOLBOIS, archéologue à l'INRAP, pour son aide dans la réflexion de création des rasters environnementaux sur QGIS et Matthieu DUCHEMIN, pour son suivi tout le long de mon stage. Je remercie aussi Amaëlle BISH, pour ses conseils concernant l'extraction de données chlorophylle, température et bathymétrie à partir de fichiers satellites, ainsi que pour sa bonne humeur qui a su me remonter le moral dans les moments difficiles.

Mes remerciements vont aussi à Hélène, Lucas et Emma, en CDD, en stage et en service civique à l'association, pour leur bonne humeur, leur aide pour la transformation des données 2016, la patience dont ils ont fait preuve pour me supporter, et tous les bons moments que j'ai passés avec eux.

Merci à Lyse SOUPLET, Présidente de l'association, et aux membres du CA, pour m'avoir donné l'opportunité de participer avec toute l'équipe d'Al Lark à la formation "Méthodologie de projet" (Formatrice : Anne Castel, de SMART (Coopérative de formation agréée UNIFORMATION)). Cela m'a permis d'éclaircir des points essentiels pour mon avenir professionnel, et de me sentir pleinement intégrée dans l'équipe.

Je souhaite également faire part de ma reconnaissance aux membres du Centre Nautique de Cancale pour leur accueil chaleureux dans leurs locaux.

J'adresse un remerciement particulier à tous les bénévoles et adhérents de l'association sans qui la collecte de données serait impossible. Merci pour votre accueil, votre participation et votre soutien.

Pour finir, merci à ma famille et à mes amis, pour avoir su me redonner le sourire et de la motivation dans cette période compliquée.

Enfin, je remercie M. Tentelier, Mme Bolliet, et l'Université de Pau et des Pays de l'Adour, pour m'avoir permis de réaliser ce stage.

## Avant-propos

### L'association

L'association AL LARK est une association loi 1901 basée à Cancale, en Ile-et-Vilaine (35). Créée en 2004 par Yohann AVICE, ancien guide nature, l'association compte maintenant plus de 2500 adhérents. Elle agit pour la sensibilisation au respect et la préservation de la biodiversité du milieu marin en baie du Mont St Michel. Dans ce but, elle a développé l'accès à la mer pour tous en proposant à ses adhérents des sorties d'observation en mer en bateau semi-rigide mais également des classes de mer, des animations pédagogiques et des actions engagées (nettoyage de plages). La baie a la particularité d'héberger une population d'environ 400 Grands Dauphins (*Tursiops truncatus*), ce qui constitue un vecteur positif de la sensibilisation.

Dans l'optique de développer les connaissances scientifiques sur les mammifères marins et donc de contribuer à une meilleure protection, lors de chaque sortie bateau, l'association collecte des données de différentes natures (données environnementales, spatiales, temporelles, comportementales, anthropiques et photographiques). Ces sorties participatives permettent ainsi le suivi scientifique de la population de cétacés (essentiellement les Grands Dauphins et les Dauphins de Risso) du golfe normand-breton.

En outre, afin de pouvoir intervenir dans le cas d'échouages de mammifères marins vivants ou morts, les salariés d'AL LARK sont correspondants pour le Réseau National Echouages et effectuent ponctuellement des nécropsies ou des prises en charge d'animaux.

### Déroulement des 6 mois de stage

Je suis arrivée en stage à Cancale le 26 février. Après deux semaines passées dans les locaux de l'association, le confinement m'a forcée à rentrer au Pays Basque chez mes parents car la connexion internet de mon appartement ne me permettait pas de travailler efficacement en télétravail (l'appartement n'étant de base pas prévu pour du télétravail, je n'avais pas d'installation internet).

Je suis restée confinée jusqu'au 6 juin, comme tout un chacun, cet environnement de travail en famille n'était pas le plus propice pour une bonne réalisation de stage ; néanmoins les réunions d'équipe tenues une fois par semaine m'ont aidées à garder mon objectif en tête. Consciente des limites de mes connaissances sur les techniques et la méthodologie de modélisation, j'avais contacté plusieurs chercheurs d'universités tierces afin de m'aider dans l'élaboration de mon stage. Néanmoins, dans le contexte particulier dans lequel nous étions, les chercheurs n'ont pas pu trouver le temps de m'accorder des rendez-vous. Cependant, j'ai pu recevoir de l'aide de la part de personnes de l'association (Jeremy, Matthieu) qui ont su m'aider à débloquer des situations et à me guider dans ma réflexion.

Je suis rentrée à Cancale le 8 juin afin de participer avec toute l'équipe à la formation de 6 jours « Méthodologie de projet », ce qui m'a permis de me remettre dans le contexte de l'association et de développer ma réflexion sur mon projet professionnel.

La taille du bureau ne permettant pas l'accueil de plus de deux personnes au sein des locaux, je suis restée en télétravail dans mon appartement du 15 juin jusqu'à la fin de mon stage, le 26 août. Je suis restée deux semaines sans internet jusqu'à l'installation de la fibre le 1er juillet, ce qui a entraîné un retard important pour le téléchargement des données externes (variables environnementales). J'ai ensuite pu travailler dans de bonnes conditions, avec des réunions avec ma maître de stage organisées toutes les semaines.

## I. Introduction

Ces cinq dernières années, AL LARK a accumulé une quantité importante de données. Un premier stage de master 1 a été effectué l'an dernier afin d'établir une gestion des données plus pertinente et organisée. Cette optimisation de collecte et de stockage des données doit faciliter leur traitement et leur exploitation pour les études développées dans ce rapport et les études à venir.

Dans le cadre de cette étude, l'intérêt se portera sur une espèce en particulier, le Grand Dauphin (*Tursiops truncatus*, Montagu 1821). Le but des recherches d'AL LARK est d'approfondir les connaissances accumulées sur la population résidente du golfe normand-breton. Car si l'espèce est très étudiée dans le monde, cette population reste encore largement méconnue. En effet, les circonstances de leur présence, de leur abondance ou de leur écologie sont encore assez vagues, ce qui laisse des possibilités importantes de recherche.

Pour une population, le choix de l'habitat est le résultat de compromis entre la disponibilité des ressources et le risque de prédation (Lima et Dill, 1990). La disponibilité des ressources est un aspect crucial qui influence la sélection des habitats (Torres *et al.*, 2008; Planque *et al.*, 2011) et façonne la répartition des prédateurs marins supérieurs tels que les cétacés, qui sont fortement influencés par les mouvements spatiaux de leurs proies (Hamazaki, 2002; Redfern *et al.*, 2006; López et Methion, 2017; 2018). Comme il est difficile d'obtenir des informations fiables sur la distribution et l'abondance des proies, les facteurs physiques et biologiques peuvent être utilisés comme des approximations pour modéliser et prévoir la distribution des prédateurs supérieurs (Guisan et Zimmermann, 2000; Redfern *et al.*, 2006; Elith et Leathwick, 2009; Pirodda *et al.*, 2011; López et Methion, 2017). De plus, dans l'environnement marin, les caractéristiques topographiques et les changements temporels des facteurs physiques et biologiques sont connus pour déterminer la répartition spatiale d'une espèce (Brodeur et Percy, 1992; Cribb *et al.*, 2015).

Comprendre l'utilisation de l'habitat et déterminer les facteurs responsables de la distribution de prédateurs marins, tel que le Grand Dauphin, est fondamental pour mieux comprendre comment ces espèces potentiellement vulnérables réagissent aux changements de leur environnement. La compréhension des relations entre les espèces et l'environnement est essentielle pour identifier les zones d'importance biologique et pour établir des priorités en matière de conservation, d'évaluation d'impact et de décisions de gestion des ressources (Guisan et Thuiller, 2005; Elith et Leathwick, 2009). De plus, malgré les informations disponibles sur les menaces qui pourraient peser sur les grands dauphins dans la zone, il y a une absence d'étude reliant la qualité de l'habitat de l'espèce à sa préservation. Dans un contexte de changement climatique global, mieux connaître les habitats clés de l'espèce dans la zone pourrait permettre de développer de meilleurs plans de conservation afin de minimiser l'impact des activités anthropiques sur l'espèce. Ces résultats seront utilisés pour évaluer quelles mesures seraient les plus pertinentes pour promouvoir la conservation de cette population, et pour évaluer les sites dans lesquels des stratégies de conservation seraient les plus efficaces.

Les modèles de distribution d'espèces (SDM) et les modèles de niche environnementale (ENM) permettent de comprendre les interactions entre un organisme et son environnement et offrent un moyen d'explorer cette réponse à différents types de facteurs environnementaux.

Ces trente dernières années, un développement statistique important a permis des avancées dans les analyses de régression, fournies par les modèles linéaires généralisés (GLM) et les modèles additifs généralisés (GAM) (Guisan *et al.*, 2002). Ces approches statistiques permettent de vérifier l'existence de relations respectivement linéaires et non linéaires entre une variable réponse et un ensemble de

variables prédicteurs. La capacité de ces outils à traiter des données non linéaires a permis de développer des modèles écologiques qui représentent mieux les données sous-jacentes, et donc d'accroître notre compréhension des systèmes écologiques (Guisan *et al.*, 2002).

Les ENM sont basées sur le concept de niche écologique (Hutchinson, 1957) qui met en relation la fitness d'une espèce à sa niche. Parmi les ENM, l'analyse des facteurs de niche écologique (ENFA) définit la niche écologique comme un hypervolume de dimension  $n$  correspondant à  $n$  variables écologiques au sein desquelles une espèce peut exister et se reproduire avec succès (Hutchinson, 1957). Cela combine des informations sur la répartition d'une espèce avec un ensemble de variables éco-géographiques (VEG) pour déterminer l'adéquation de l'habitat (Hirzel *et al.*, 2002).

Les informations sur la répartition d'une espèce peuvent être enregistrées en termes de données de présence/absence (Weir *et al.*, 2012; López et Methion, 2017; 2018) ou de données de présence uniquement (Moura *et al.*, 2012; Fernandez *et al.*, 2018). Cependant, la détection d'espèces marines très mobiles, telles que le grand dauphin peut être difficile car elles passent de courtes périodes à la surface (Hamazaki, 2002; Praca et Gannier, 2008). Par conséquent, la distinction entre les vraies absences (c'est-à-dire lorsque les dauphins ne sont pas présents dans l'endroit échantillonné) et les fausses absences (c'est-à-dire lorsque les dauphins sont présents mais n'ont pas pu être détectés) peut être difficile (Praca et Gannier, 2008 ; Elith et Leathwick, 2009). Plus précisément, il peut être difficile d'évaluer si ce petit cétacé est absent dans un endroit connu parce que (1) l'habitat n'est pas adapté aux grands dauphins, (2) l'habitat est adapté, mais les dauphins ne sont pas présents au moment où passe le bateau, ou (3) les dauphins sont présents mais n'ont pas pu être détectés.

La différence majeure entre les trois modèles de distribution des espèces décrits précédemment est la qualité des données de présence/absence nécessaires. Les méthodes telles que les GLM et les GAM requièrent des données de présence/absence de bonne qualité afin de générer des fonctions statistiques ou des règles discriminantes qui permettent de classer la probabilité d'habitat en fonction de la répartition de la présence et de l'absence des espèces (Manel *et al.*, 1999; Guisan et Zimmermann, 2000).

La méthode ENFA requiert seulement des données de présence et a été développée pour permettre l'utilisation de données où les absences sont indisponibles ou inadaptées (Carpenter *et al.*, 1993; Hirzel et Guisan, 2002; Farber et Kadmon, 2003). Ces modèles peuvent atteindre une grande précision prédictive avec des échantillons de petite taille (Allouche *et al.*, 2008), et sont idéaux pour évaluer l'adéquation de l'habitat des espèces très mobiles (Reutter *et al.*, 2003; Pracca et Gannier, 2008). Dans notre cas, ce type de modèle est donc recommandé (Hirzel *et al.*, 2002 ; MacLeod *et al.*, 2008), et s'avère être une technique robuste pour évaluer l'adéquation de l'habitat des cétacés (Praca et Gannier, 2008 ; Skov *et al.*, 2008; Condet et Dulau-Drouot, 2016). En outre, ils sont utilisés pour déduire les menaces potentielles à la conservation des prédateurs marins, telles que la perte d'habitat ou l'interaction avec les activités humaines (Condet et Dulau-Drouot, 2016). De plus, utilisant une espèce virtuelle avec des préférences prédéfinies en matière de sélection d'habitat, Hirzel et collaborateurs (2001) ont comparé les performances de modèles issus d'une méthode reposant uniquement sur la présence des espèces (ENFA) avec une méthode qui nécessite à la fois des données de présence et d'absence (GLM). Bien que les deux méthodes aient fourni de bonnes prédictions de la distribution virtuelle des espèces, les auteurs ont constaté que l'ENFA avait tendance à être plus performante dans les scénarii où les espèces n'occupaient pas tout l'habitat adapté (c'est-à-dire que de nombreuses absences étaient ainsi modélisées dans l'habitat adapté). À l'inverse, lorsque les espèces étaient modélisées pour utiliser tous les habitats optimaux avec une probabilité élevée et modélisées pour utiliser des habitats inadaptés avec des probabilités plus faibles, les GLM étaient plus précis. Étant donné que la qualité des données est susceptible d'être une question clé affectant la

fiabilité des prévisions des modèles (Zaniewski *et al.*, 2002; Stockwell et Peterson, 2002), la connaissance des performances prédictives des méthodes et de leur domaine d'application devient une question importante aux premiers stades de l'élaboration des projets visant à cartographier la répartition des espèces. À l'heure actuelle, il existe un manque de tests approfondis sur les performances relatives des méthodes qui comparent les modèles de distribution des espèces en utilisant des données de présence seule ou de présence/absence. Si l'utilisation d'espèces virtuelles est utile pour l'exploration préliminaire du comportement des méthodes, il est important d'utiliser des données réelles sur la distribution des espèces pour mieux comprendre les performances relatives des méthodes. En outre, étant donné qu'il est difficile d'obtenir des données précises sur les absences, notamment pour les espèces mobiles ou peu visibles, il est particulièrement important d'étudier les circonstances qui font que les modèles utilisant des données de présence uniquement sont au moins aussi performants que les modèles utilisant des données de présence/absence. Cela devrait permettre de mieux comprendre les méthodes qui sont plus adaptées à des applications particulières.

Il est également intéressant de noter qu'en raison de la grande variété de méthodes, de types de données et d'algorithmes de modélisation disponibles pour les SDM, il existe des contradictions entre les résultats des techniques à modèle individuel. En effet, les modèles les plus performants tendent à varier selon la zone ou l'espèce. Une méthode permettant de surmonter ces variations consiste à combiner les prévisions d'un modèle unique par un processus connu sous le nom de modélisation d'ensemble, ou consensus (Araújo et New, 2007 ; Franklin, 2010) La modélisation d'ensemble fournit souvent des estimations plus robustes des distributions des espèces car les prévisions des modèles combinés donnent des précisions plus élevées et moins de biais que les modèles individuels séparés (Marmion *et al.*, 2009; Grenouillet *et al.*, 2011; Scales *et al.*, 2016). Si la modélisation d'ensemble a déjà été utilisée pour certaines espèces marines (Riul P *et al.*, 2013; Pikesley *et al.*, 2015), elle ne semble pas encore largement utilisée pour modéliser la distribution des dauphins (mais voir Moura *et al.*, 2012; Pérez-Jorge *et al.*, 2015).

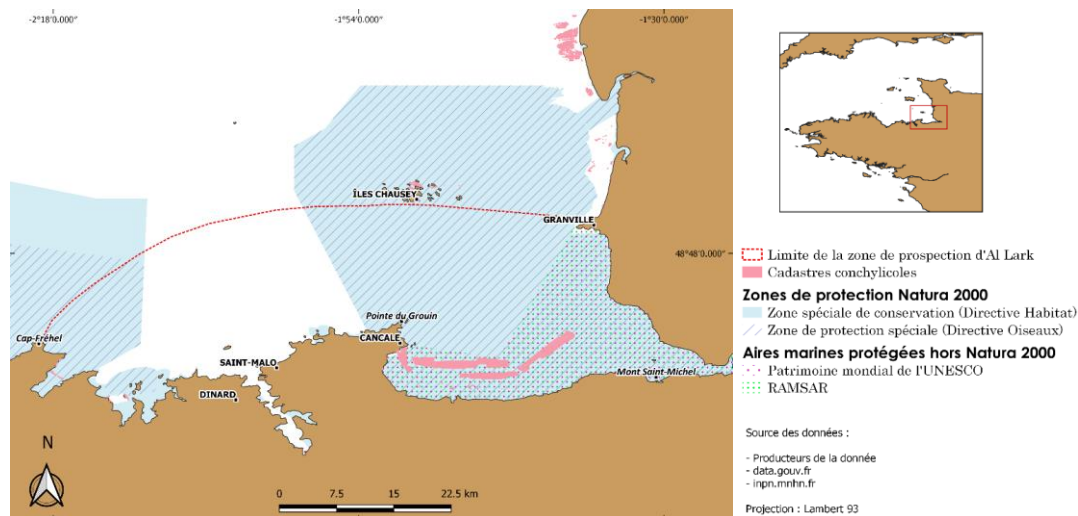
Cette étude combine la modélisation de la distribution des espèces (GLM, GAM, ENFA) avec des données présentielle collectées au cours des sorties bateaux afin d'obtenir des informations sur les variables environnementales qui pourraient influencer la distribution et l'habitat des grands dauphins. Ici, les habitats préférés sont identifiés comme des zones de forte probabilité de présence de dauphins, ce qui reflète leur distribution dans l'espace et le temps (Hastie *et al.*, 2004).

L'objectif de ce travail est (1) d'étudier la distribution du grand dauphin en fonction de certaines variables environnementales, (2) d'identifier tout habitat préférentiel et (3) de comparer les performances de prédiction de plusieurs techniques de modélisation différentes : le modèle linéaire généralisé (GLM), le modèle additif généralisé (GAM), la modélisation d'ensemble et l'ENFA (Ecological Niche Factor Analysis).

## II. Matériel et méthodes

### 1. Site d'étude

Le golfe normand-breton est situé en Manche occidentale, il possède un littoral de 450 km<sup>2</sup> et couvre une superficie totale de 14 000 km<sup>2</sup> (figure 1). Il est ponctué de nombreuses îles qui influencent la circulation et la direction des courants de marée. La bathymétrie dans le golfe n'exécède pas les 60 m de profondeur (Trigui, 2009). Située au fond du golfe normand-breton, la baie du Mont-Saint-Michel est une enclave sédimentaire de près de 500 km<sup>2</sup> s'étendant au sud d'une ligne qui joint la pointe du Grouin à Cancale à l'ouest à la pointe du Roc à Granville à l'est (Bonnot-courtois, 2012).



La baie est caractérisée par de forts gradients hydrodynamiques qui génèrent une grande diversité de faciès sédimentaires (Ehrhold, 1999). En ce qui concerne la géomorphologie de la zone, la baie est composée d'immenses estrans calcaro-sableux se terminant par des schorres et des polders. Le mont se trouve à 16 milles des fonds de -20 mètres. Entre la Pointe du Grouin et le Cap Fréhel, se trouve la baie dite Malouine, composée de granites, migmatites, gneiss et schistes.

Les apports d'eau douce sont faibles voire négligeables dans l'ensemble de la zone (0.5‰) et proviennent principalement de quelques fleuves côtiers : Le Couesnon, La Sée et La Sélune dans la baie, et la Rance à Saint-Malo (Bonnot-courtois, 2012). La salinité est en général inférieure à 35 ‰ excepté à l'automne où elle peut être légèrement supérieure (Doums *et al.*, 2013).

Dans la baie du Mont Saint Michel, le courant tourne en sens inverse des aiguilles d'une montre; des valeurs maximales de 3 nœuds ont été signalées<sup>1</sup>. Le marnage dans la baie peut aller jusqu'à 15,5 m, découvrant, lors des grands coefficients de marée, une vaste zone d'estran de près de 240 km<sup>2</sup> (Le Mao *et al.*, 2006). Les faibles profondeurs associées aux courants de marée intenses provoquent un brassage continu des eaux entraînant de ce fait une forte turbidité (8-17 mg/l) et une faible différence de température entre la surface et le fond (Doums *et al.*, 2013). Les niveaux de turbidité présentent néanmoins un signal saisonnier, avec des valeurs maximales en période hivernale, de décembre à février, et des valeurs minimales en période estivale, de juin à août (Gohin, 2010). La température moyenne hivernale va de 8,5°C au large à 7,5°C au fond du golfe et la température moyenne estivale va de 16,5°C à 17°C. Dans la baie, près du zéro hydrographique, les températures



peuvent atteindre 22°C du fait de l'extension des estrans. Les températures constituent de ce fait le principal facteur de caractérisation des masses d'eau<sup>1</sup>. À l'échelle du golfe, la température des eaux de surface fluctue selon un gradient d'orientation nord-ouest / sud-est qui s'inverse en fonction des saisons. Ainsi, en été, les eaux les plus chaudes sont enregistrées au fond de la baie du Mont Saint Michel alors que des eaux plus froides sont enregistrées au nord du golfe, et inversement en hiver (Trigui, 2009). En réponse au changement climatique récent, les eaux de la Manche, et plus spécifiquement celles du Golfe, se sont réchauffées de manière significative au cours des trente dernières années, avec de possibles répercussions sur la distribution de la macrofaune benthique (Ménèsquen *et al.*, 2007 ; Gaudin, 2017).

En outre, Ménèsquen et ses collaborateurs (2007) ont démontré que la distribution de la biomasse phytoplanctonique annuelle moyenne évolue selon un gradient spatial Est/Ouest. En effet les valeurs maximales ( $> 1,5 \text{ mg.m}^{-3}$ ) sont observées le long du littoral du Cotentin jusqu'à la baie du Mont-Saint-Michel. Les valeurs minimales ( $< 0,75 \text{ mg.m}^{-3}$ ) sont rapportées dans les eaux du large, à l'ouest d'une ligne reliant Aurigny, Jersey et le cap Fréhel (Ménèsquen *et al.*, 2007; Gohin, 2010). De plus, il existe aussi une variation spatiale des périodes de blooms phytoplanctoniques, les concentrations en chlorophylle-*a* sont ainsi maximales en baie du Mont-Saint-Michel fin avril-début mai, alors qu'elles ne le sont qu'en juillet-août à l'ouest de Jersey et Guernesey (Ménèsquen *et al.*, 2007).

En ce qui concerne la production primaire brute, qui donne des informations sur le fonctionnement des écosystèmes pélagiques, elle montre un schéma spatial différent de la distribution de la biomasse chlorophyllienne. En effet, les valeurs minimales ( $< 25 \text{ gN.m}^{-2}.\text{an}^{-1}$ ) ont été estimées en particulier le long du Cotentin, et les valeurs maximales ( $< 40 \text{ gN.m}^{-2}.\text{an}^{-1}$ ) à l'ouest du golfe (Ménèsquen et Hoch, 1997). Malgré des apports nutritifs plus importants à la côte, la production primaire est limitée par la faible profondeur et la forte turbidité des eaux. Néanmoins selon Laffaille et ses collaborateurs (2000), en baie du Mont Saint Michel, la forte turbidité sur les vastes zones d'estran pourrait favoriser la production primaire microphytobenthique, estimée à  $23.5 \text{ gC.m}^{-2}.\text{an}^{-1}$  par Migné et ses collaborateurs en 2009 (Migné *et al.*, 2009).

On peut également noter que le golfe normand-breton subit une forte pression anthropique sur son littoral, due à la pêche, à la conchyliculture et au tourisme.

La zone de prospection de l'association couvre une partie de la Côte d'Emeraude. Cette zone est délimitée à l'ouest par l'embouchure de la Rance, à l'Est par Granville, au nord par les îles Chausey et couvre toute la baie du Mont Saint Michel (Figure 1). Elle est divisée en deux parties : l'intérieur et l'extérieur de la baie. La surface de la zone d'étude est d'environ  $1000 \text{ km}^2$ .

---

1

<https://envlit.ifremer.fr/region/bretagne/milieu/hydrologie#:~:text=Ille%2Det%2DVilaine%20et%20C%3%B4tes%20d'Armor&text=La%20salinit%C3%A9%20est%20en%20g%C3%A9n%C3%A9ral,C%20C%3%A0%2017%C2%BC>.

## 2. Le Grand Dauphin *Tursiops truncatus*

- Présentation de l'espèce

Le Grand Dauphin (*Tursiops truncatus*, Montagu 1821)(Ross, 1977), est une espèce de mammifère marin de l'ordre Cetacea et de la famille des Delphinidae (Ross, 1977). Il est classé comme Least Concern (Préoccupation mineure) sur la liste des espèces menacées de l'IUCN (iucnredlist.org) et est



Figure 2. Photographies de *Tursiops truncatus*

une espèce prioritaire pour la conservation dans les eaux européennes car listé dans les annexes II et IV de la directive Habitat 92/43/CEE de l'Union Européenne.

Cette espèce (Figure 2) est cosmopolite (Perrin *et al.*, 2009) et bien qu'elle occupe surtout les eaux côtières, comprenant les estuaires, les baies et l'aval de certaines rivières, elle peut aussi fréquenter les eaux pélagiques (Wells et Scott, 1999). L'utilisation de l'habitat, pour la profondeur, est extrêmement variable d'un endroit à l'autre (Natoli *et al.*, 2005;

Würsig, 1978). La température de surface est aussi un paramètre important mis en évidence dans différentes études, comme celle de Hastie *et al.* (2005), qui montre une relation entre la distribution du grand dauphin avec une température comprise entre 8 et 12°C. Aussi, la concentration de chlorophylle à la surface fait partie des variables typiques pouvant influencer la distribution d'une population de cétacés (Smith *et al.*, 2008; Davis *et al.*, 2002). De nombreuses études ont montré que la distribution des cétacés est liée à des facteurs environnementaux, à la fois physiographiques (comme la profondeur, le substrat) et océanographiques (par exemple la température, la salinité, ou la concentration en chlorophylle-a) à différentes échelles. Ces interactions peuvent être directes ou indirectes (Spyrakos *et al.*, 2011).

Le régime alimentaire du *Tursiops truncatus* est principalement composé de poissons et de céphalopodes (Santos *et al.*, 2007). Cette espèce est donc considérée comme un prédateur supérieur (Wells *et al.*, 2004). De plus, avec sa longue espérance de vie et sa capacité à résider à long terme dans le même lieu, elle peut être présentée comme un bio-indicateur de l'état des écosystèmes marins côtiers (Tabor et Aguirre, 2004; Wells *et al.*, 2004).

- Présence dans le golfe normand-breton

Dans les eaux côtières de la Manche et de l'Atlantique, trois populations de Grands Dauphins ont été répertoriées : deux petites (dizaine d'individus) en mer d'Iroise et une dans le golfe normand-breton (Louis *et al.*, 2014) Cette population se compose d'environ 400 individus sédentaires (IC 95% : 290-380) (Gally, 2016). Des études (Louis *et al.*, 2015 ; Blandel, 2018) ont montré que la population serait divisée en trois sous-groupes interconnectés (Annexe 3) : un dans la partie nord du golfe (—North), un autre au centre (—Miquiers) avec des individus utilisant l'ensemble de la zone et un troisième dans la partie sud du golfe (—South). L'association AL LARK couvre géographiquement la zone correspondant majoritairement aux individus de la sous-population « Sud » (Perri et Petiau, 2019)

### 3. Les données collectées

#### 3.1. Généralités

L'association AL LARK organise des sorties en mer de mars à novembre, lorsque le nombre d'adhérents est suffisant (8 minimum) et que les conditions météorologiques sont optimales (vent inférieur à 4 Beaufort, bonne visibilité sans brume ni pluie, mer peu ou pas formée). Le départ des sorties se fait au niveau de la plage de Port-Mer à Cancale. La fréquence des sorties varie donc de quelques sorties par semaine à plusieurs par jour selon les conditions météorologiques. Les routes empruntées varient d'une sortie à l'autre afin d'obtenir une couverture maximale et représentative de la zone d'étude et de s'assurer qu'aucun secteur potentiellement important n'ait été négligé, néanmoins la zone de prospection est souvent influencée par l'état de la mer. Durant l'effort de prospection, le bateau navigue à une vitesse d'environ 10 à 12 nœuds, et les observateurs scrutent (à l'œil nu et aux jumelles) une zone d'environ 360 degrés autour de l'axe de la plateforme d'observation dans le but de détecter la présence de cétacés. Des pauses régulières sont effectuées tout au long de la sortie afin d'optimiser la recherche et de limiter la fatigue visuelle liée aux jumelles. Les données environnementales (météo, état de la mer) sont collectées au début de la sortie et à chaque changement lors de la sortie. Lors de la détection d'un groupe ou d'un individu, le protocole d'approche est d'abord mis en place, selon les critères du label High Quality Whale Watching du sanctuaire PELAGOS (PELAGOS, 2007), puis le protocole de collecte des données sur les grands dauphins est mis en place. Sur les quatre années de suivi étudiées ici, deux protocoles différents ont été utilisés. Les supports utilisés sont l'application TapForm pour 2016 et 2017, et l'application OBSenMER pour 2018 et 2019 (Annexe 5).

#### 3.2. Données collectées par suivis scientifiques

##### - Données environnementales

Selon les différents protocoles mis en œuvre, les données environnementales collectées ne sont pas toujours les mêmes. Le protocole TapForm collecte certaines données qui ne peuvent pas être collectées avec OBSenMer (informations sur la marée) et inversement (état de la mer). De plus, certains paramètres environnementaux sont pris en compte par les deux protocoles, mais sous des unités différentes (Annexe 4). Un des objectifs du stage de master 1 a été de créer une base de données permettant de centraliser et d'homogénéiser l'ensemble des données scientifiques acquises par ces différents protocoles pour les années 2017 et 2018 (Annexe 7). Il a donc fallu aussi pour cette étude continuer ce travail, et aller chercher les informations manquantes des protocoles 2016 et 2019 pour continuer à incrémenter la base de données. Pour chaque sortie, les valeurs de coefficient de marée et d'état de la marée ont été rajoutées à partir des horaires des marées du Service hydrographique et océanographique de la Marine (SHOM) pour le port de Cancale (d'après les données officielles : [www.maree.info](http://www.maree.info)). Certaines données manquantes pour plusieurs sorties (données météo notamment) ont été récupérées à partir de la base de données du site OpenWeatherMap (<https://openweathermap.org/>). Les variables temporelles (saison, moment de la journée, mois) ont aussi été créées *a posteriori* à partir des trackings gps des sorties. De nouvelles variables ont été rajoutées en fonction des observations sur le terrain et/ou de manière à apporter de la matière pour de futures analyses scientifiques, comme par exemple le temps de pause, l'état de la mer en Beaufort, la médiane du nombre d'individus, le moment de la journée, le code de photo-identification, et l'indice de présence.

##### • Trajet

Le tracé GPS correspondant au trajet effectué par le bateau est enregistré pour chaque sortie. Pour le protocole utilisé en 2016 et 2017, les tracés GPS ont été enregistrés sous une application tierce

(Motionx gps HD). Ils sont seulement récupérables au format .gpx, donc directement lisible sous un SIG, mais non modifiables. Pour les années 2018 et 2019, les tracés GPS sont extraits à partir d'OBSenMER sous un format .csv, néanmoins, les données 2018 ont aussi nécessité un reformatage pour permettre la lecture sous SIG.

- Présence de Grands Dauphins

Lors de la détection de grands dauphins, un certain nombre de paramètres sont enregistrés, pour les deux types de protocoles. Pour le protocole 2016-2017, l'heure de rencontre et la position GPS du groupe devaient être notés manuellement. Si le groupe avait pu être suivi pendant quelques temps, l'heure et les coordonnées de fin de suivi pouvaient être notées de la même manière. Avec OBSenMER, les informations spatio-temporelles de la rencontre sont enregistrées automatiquement dès la notification d'une observation. Dans le cas d'un suivi, l'outil de tracé GPS compris dans OBSenMER enregistre toutes les minutes des points GPS pouvant être utilisés pour visualiser le trajet du groupe, et ce jusqu'à ce que le chef de mission arrête ce suivi.

D'autres données (taille de groupe, présence de jeunes, comportement), destinées à des études de population et d'éthologie, sont aussi récoltées dans les deux protocoles (Annexe 6)

### 3.3. Données supplémentaires

- Données de distribution de l'espèce

La mise en place d'un modèle prédictif d'habitat repose avant toute chose, quel que soit le modèle, sur les points géographiques d'observations de l'espèce étudiée. Les occurrences doivent pouvoir être géo-référencées et insérées dans un SIG. Le plus simple est de travailler sans projection, et d'utiliser toujours le même système de coordonnées géographiques WGS84 pour le domaine maritime. Selon la méthode utilisée (ENFA sous Biomapper4, ou GLM/GAM sous R) le format des fichiers demandés, contenant les données de distribution connues de l'espèce considérée, est différent. Ainsi :

- pour l'ENFA il faut un raster contenant la valeur 1 dans les cellules où l'espèce est présente et 0 dans les cellules dont on ne sait rien (raster booléen) ;
- pour les GLM et les GAM, qui tiennent compte aussi des points d'absence, il faut un fichier \*.txt avec trois ou quatre colonnes séparées par des virgules (id sans entête, Latitude, Longitude, Présence) contenant la liste des points de présence et des points d'absence. La colonne Présence sera remplie de 1 et de 0 pour distinguer les enregistrements de présence ou d'absence.

Le protocole d'échantillonnage utilisé lors des sorties ne nous permet pas d'avoir des données d'absence/présence, seule la présence ponctuelle ou la notification d'un suivi est enregistrée sur la tablette. Pour pallier cela, un recodage des données est nécessaire. Le fichier contenant les observations géoréférencées et le fichier contenant les tracking sont recoupés manuellement afin d'obtenir des 0 à tous les points de tracking où aucune observation n'a été enregistrée, et des 1 seront notés sur les points géoréférencés où une observation ponctuelle ou un suivi ont été enregistrés. Il faut noter que l'intervalle de temps de prise de point GPS n'est pas régulier (il varie de 30s à 2min). A partir du fichier complet contenant les absences/présences, un tri a été effectué afin d'obtenir des points régulièrement répartis le long de la ligne de transect. De plus, un point d'absence et un point de présence ne doivent pas se trouver dans une même maille de la grille. Et afin d'éviter toute confusion avec les données de présence et ne pas avoir un point absence et un point présence dans la même cellule, certains points d'absence ont été supprimés.

Les effectifs de Présence/Absence ont ensuite été équilibrés afin d'être égaux pour ne pas biaiser les résultats finaux. Les données d'absence comportent potentiellement un biais du fait du traitement effectué, on comparera donc des modèles issus de nos données d'absence avec des modèles utilisant des pseudo-absences générées automatiquement par les analyses GLM et GAM (package BIOMOD2) .

- Données environnementales issues de rasters mensuels

Afin d'étudier la distribution des grands dauphins sur notre site d'étude, plusieurs variables environnementales ont été récupérées sur des supports externes à l'association. Les logiciels QGIS 2.10 (QGIS Development Team, 2010) et RStudio, ont été utilisés pour extraire les valeurs des variables environnementales à chaque échantillon de 10 min de présence/absence. Les valeurs de salinité (unité : mg/m<sup>3</sup>) ont été extraites à partir d'un raster NetCDF-4 mensuel<sup>2</sup> issu du site web COPERNICUS Marine Environment Monitoring Services (<http://marine.copernicus.eu>), de résolution 0.111 x 0.067 degré. L'extraction et la transformation de ce raster au format GeoTIFF ont été effectuées grâce au package NETCDF (QGIS 2.10). La température de surface de l'eau a été obtenue à partir de rasters satellitaires mensuels de résolution 0.1 degré du site web NASA Earth Observation (<https://neo.sci.gsfc.nasa.gov/>)<sup>3</sup>. Après extraction des variables pour chaque point, un changement d'unité a été nécessaire pour obtenir les valeurs de température en degré Celsius. Les concentrations en Chlorophylle-a (mg/m<sup>3</sup>) ont été extraites à partir de rasters mensuels de résolution 0.1 degré du site NASA Earth Observation<sup>4</sup>.

La variable « Type de substrat » a été récoltée à partir d'une carte de nature du fond au 1/150000 du SHOM<sup>5</sup> au format shapefile. Afin de rasteriser cette carte, les valeurs de type de substrat ont été transformées en valeurs quantitatives (VOIR TABLEAU ANNEXES). Les outils « PointSamplingTool » et « Joindre les attributs par localisation » de QGIS 3.10 ont été utilisés pour extraire les valeurs de substrat à chaque point de présence/absence.

- Calcul de Nouvelles variables

Les données de bathymétrie (profondeur du fond en mètre à partir du zéro hydrographique, c'est-à-dire le niveau théoriquement atteint par les plus basses mers astronomiques [coefficient 120]) sont issues de la Carte Bathymétrique Générale des Océans (GEBCO20, <http://www.gebco.net>).

Le gradient et l'aspect de la pente sont calculés à partir de la profondeur du fond obtenue à partir des données de la carte bathymétrique du GEBCO grâce aux packages « rgdal » et « raster » de R.

Deux autres indices de terrain sont aussi calculés selon Wilson et al. (2008) :

- le TRI (Terrain Ruggedness Index), moyenne des différences absolues entre la valeur d'une cellule et la valeur des 8 cellules qui l'entourent
- la rugosité, définie comme l'écart-type de la profondeur dans chaque pixel. Cet écart-type reflète la rugosité du fond et donne une image du relief sous-marin.

La hauteur de la marée a été obtenue à partir des données du marégraphe de Saint Malo (<https://data.shom.fr/>).

---

<sup>2</sup>[https://resources.marine.copernicus.eu/?option=com\\_csw&view=details&product\\_id=NORTHWESTSHELF\\_REANALYSIS\\_PHY\\_004\\_009](https://resources.marine.copernicus.eu/?option=com_csw&view=details&product_id=NORTHWESTSHELF_REANALYSIS_PHY_004_009)

<sup>3</sup> <https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MYD28M&year=2016>

<sup>4</sup> [https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MY1DMM\\_CHLORA&year=2017](https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MY1DMM_CHLORA&year=2017)

<sup>5</sup> <https://diffusion.shom.fr/pro/ressources/sedimentologie/carte-sedimentologique-au-1-150-000.html>

La baie du Mont Saint Michel présentant un marnage pouvant atteindre 15 mètres, les données bathymétriques peuvent ne pas refléter la hauteur d'eau réelle sous le bateau au moment de l'observation. Celle-ci (appelée par la suite profondeur (en m)) a été calculée en couplant les données de bathymétrie issues du GEBCO, et des données de hauteur de marée du marégraphe. En effet, lorsque les points de présence/absence disposaient de valeurs positives de bathymétrie, la valeur retenue était seulement la hauteur de la marée. Pour les points où les valeurs de bathymétrie étaient négatives (observation au-dessus du zéro hydrographique, donc sur l'estran) (Annexe 8), on a additionné la hauteur d'eau à la valeur de bathymétrie.

Une variable « Distance à la Côte » a été calculée grâce au plugin NNjoin de QGIS 2.18 en mesurant la distance minimale en kilomètres entre les coordonnées GPS de chaque point et la côte de la zone d'étude. L'isobathe 0, ou zéro hydrographique, peut être utilisé pour déterminer la côte dans certains cas, néanmoins, à cause du grand marnage observé dans la zone d'étude, cet isobathe peut se situer jusqu'à 13 kilomètres de la côte, c'est pourquoi il n'a pas été retenu pour cette variable.

#### 4. Analyses préliminaires sur la période 2016-2019

La procédure de traitement des données a été effectuée sous QGIS Desktop 3.4. LTR « Madeira » et Excel. Les cartes ont été réalisées sous QGIS et les analyses statistiques ont été réalisées à l'aide du logiciel RStudio, version 3.5.1. Tous les tests statistiques ont été réalisés avec un  $\alpha = 5\%$ .

- Effort de prospection

Une carte de répartition de l'effort par maille a été réalisée par année. Six classes correspondant à l'effort effectué en kilomètres ont été définies. Les étapes de la réalisation de cette carte sont les suivantes :

a. Les analyses sont faites dans une grille hexagonale régulière couvrant l'ensemble de la zone d'étude. Cette méthode de spatialisation des données permet d'homogénéiser les surfaces d'analyse, donc de comparer les résultats entre eux. La grille hexagonale a été choisie pour trois raisons : (1) La zone visuelle depuis le navire de recherche est de forme circulaire, ainsi, la forme hexagonale offre une meilleure représentation que les cellules rectangulaires ou triangulaires (Paradell *et al.*, 2019); (2) De par leur forme, les cellules hexagonales ont un rapport périmètre-surface plus proche d'un cercle, ce qui pourrait réduire l'effet de bordure (Birch *et al.*, 2007); (3) Les cellules voisines sont toutes à égale distance, donc les centroïdes des cellules adjacentes sont à la même distance (pour rasterisation) (Birch *et al.*, 2007). Une surface de maille de 2x2 km a été choisie car, avec une vitesse moyenne de 10 nds, deux échantillons consécutifs de 10 min seraient situés dans des cellules adjacentes, de plus cette taille s'adapte aussi à l'espace visuel du navire. La grille est en EPSG 2154 RGF93 / Lambert-93 est a été réalisée avec le plugin MMQGIS.

b. Des polygones (Annexe 9) sont créées en reliant entre eux les différents points de chaque sortie en effort. Cette manipulation se fait à l'aide de l'outil "Points vers lignes" intégré à QGIS. Le nombre de segments en effort s'élève à 321 sur l'ensemble de la période 2017-2018.

c. La couche « polyligne » est ensuite intersectée avec la grille (Outil Somme des longueurs des lignes) et les distances kilométriques de chaque segment mesurées (fonction \$length). Cette procédure est répétée pour chaque mois, et pour chaque moment de la journée. On obtient ainsi un effort global, mensuel et circadien, par maille (en km).

L'effort calculé sous forme de distance permet de minimiser les biais liés aux changements de vitesse des différents bateaux. La différence d'effort entre ces périodes a été testée par des analyses ANOVA à un facteur sur le logiciel de statistique R, les résidus respectant les conditions d'homoscédasticité, de normalité et d'indépendance conditionnant l'utilisation de ce test.

- Distribution spatio-temporelle

Afin de représenter la distribution spatiale de l'espèce dans la zone d'étude, différentes analyses ont été réalisées. Deux indices ont été calculés pour chaque maille de la zone :

- Le taux de rencontre (TR), exprimé en nombre d'observations par kilomètre (km) prospecté, est défini tel que :  $TR = No/E$  avec No, le nombre d'observation de l'espèce et E, l'effort en km.
- L'abondance relative (AR), exprimée en nombre d'individus par kilomètre (km) prospecté, est définie telle que :  $AR=Ni/E$  avec Ni, le nombre d'individus observés et E, l'effort en km.

Ces deux indices ont été calculés pour la période 2016-2019. Des tests ANOVA ont été utilisés pour étudier les variations du taux de rencontre et de l'AR en fonction des années. Pour cartographier cet indice, les observations de cétacés sont additionnées par maille grâce à l'outil "Nombre de points" dans polygones de QGIS, puis divisées par l'effort de prospection. Le même outil est utilisé pour connaître le nombre d'individus par maille, en pondérant le nombre de points par le poids de la variable "nombre d'individus". Le calcul décrit plus haut est appliqué pour obtenir les abondances relatives.

Quel que soit l'effort effectué dans une maille donnée, le taux de rencontre et l'abondance relative ont été calculés. De ce fait, on peut obtenir un indice anormalement élevé lorsque des cétacés ont été observés pour un effort minime. Il est estimé que 10% de la surface d'une maille doit être prospecté pour que ces indices soient considérés comme représentatifs (Gannier, 1995). Dans le cas de cette étude, la surface de la maille est d'environ 2 km<sup>2</sup>. La largeur totale de détection des grands dauphins étant estimée à 1 km (David et Di-Méglio, 1999) il faut prospecter au minimum 200 m pour échantillonner 10 % de la maille. De cette manière, une maille non prospectée ou dans laquelle l'effort est inférieur à 0,2 km sera considérée comme non validée et retirée de la carte.

## 5. Analyse de la distribution spatiale des Grands Dauphins

Pour étudier la distribution des Grands Dauphins dans le Golfe Normand Breton, plusieurs variables environnementales ont été choisies et divisées en deux catégories (variables persistantes et non persistantes) en fonction de leur variabilité temporelle (López et Methion, 2017) (Tableau 1).

Variables persistantes (pas de variation temporelle)	Variables non persistantes (variations temporelles)
- <b>Distance à la côte (km)</b>	- <b>Température de l'eau de surface (°C)</b>
- <b>Bathymétrie (m)</b>	- <b>Concentration en chlorophylle-a (mg/m<sup>3</sup>)</b>
- <b>Aspect</b>	- <b>Salinité (‰)</b>
- <b>Gradient de la pente</b>	- <b>Hauteur de marée (m)</b>
- <b>Indice TRI</b>	- <b>Profondeur (m)</b>
- <b>Rugosité du fond</b>	
- <b>Type de substrat</b>	

Tableau 1: Liste des variables utilisées dans l'analyse spatiale et réparties selon leur persistance

### 5.1. Préparation des données

- Extraction des données des rasters téléchargés

Une fois les données extraites de chaque raster environnemental comme décrit plus haut, les données environnementales ont été fusionnées sur un même fichier .csv grâce au logiciel R. Ensuite toutes les données manquantes (NA) et extrêmes ont été supprimées du jeu de données. Ainsi, on obtient un fichier avec des points d'absence/présence et les valeurs environnementales associées. Une analyse préliminaire (statistiques descriptives) a été effectuée pour étudier nos données de manière générale. Afin de se rendre compte de la distribution des données et de leurs plages de valeurs, les statistiques de base<sup>6</sup> ont été calculées et des boîtes à moustaches et histogrammes ont été réalisés pour chaque variable. Afin d'étudier s'il existe un lien entre deux variables, le coefficient de corrélation de Spearman a été utilisé. Ce coefficient est une corrélation de rang qui examine s'il existe une relation entre le rang des observations pour deux variables, ce qui permet de détecter l'existence de relations monotones (croissante ou décroissante), quelle que soit leur forme précise (linéaire, exponentielle, puissance, ...). Ce coefficient est donc très utile lorsque l'analyse du nuage de point révèle une forme curviligne dans une relation qui semble mal s'ajuster à une droite ou lorsque les distributions des variables sont dissymétriques et/ou comportent des valeurs exceptionnelles. La matrice des corrélations a été utilisée dans une Analyse en Composante Principale (ACP) afin de résumer l'information dans un espace de faible dimension (au lieu des 12 dimensions apportées par les 12 variables) tout en déformant le moins possible la réalité et en analysant la dispersion des données. Cela nous donne une idée des variables potentielles que l'on pourra garder dans nos analyses.

- Création d'un raster booléen pour l'ENFA

Pour montrer les zones où la présence de dauphins a été enregistrée, un raster booléen a été créé. Ce traitement a été réalisé grâce au logiciel à QGIS 3.10. En premier lieu, la grille hexagonale de 2x2 km couvrant la zone d'étude a été utilisée (Coordonnées -2°17.45106571153627'N - 48°5771909302842957'W, -1°33.73527667904048'N - 48°96.71909302842963'W). La grille est composée de 912 cellules et est en projection EPSG:4326 - WGS 84. Le fichier de points de présence a ensuite été intégré au SIG. Avec la couche de points de présence et grâce à l'outil « Nombre de points dans les polygones », on compte le nombre de point de présence dans chaque hexagone, enregistré sous le nom « NBPT » dans la table attributaire. Dans celle-ci, une nouvelle variable « Présence » est ensuite créée avec la formule « IF NBPT>0,1,2 », ce qui code un 1 pour chaque cellule comportant au minimum un point de présence et un 2 pour les autres cellules. On notera que la formule aurait pu être directement utilisée pour coder les absences en 0, néanmoins, lors de la rasterisation, le logiciel reconnaît alors les 0 comme des NoData. Pour remédier à ce problème, les cellules contenant les points d'absence sont d'abord codés en 2, la couche shapefile est ensuite rasterisée, puis l'outil « Calculatrice Raster » est utilisé pour reclasser les cellules notées comme 2 en 0.

Ainsi, on obtient notre raster booléen, dont les cellules contenant les observations de grands dauphins sont notées 1 et celles où la présence de l'espèce n'a pas pu être prouvée, notées 0.

- Traitement des données environnementales

Pour réaliser une ENFA sur Biomapper, les données d'occurrence de l'espèce mais aussi les variables environnementales doivent être transformées au format raster (Hirzel *et al.*, 2002). On notera que les 12 cartes VEGs et le raster booléen ont été créés avec la même projection, la même

---

<sup>6</sup> Valeurs minimum, maximum, moyennes, écart-types, variances, médiane, quantiles



emprise, la même taille de pixel (0.002), et la même taille de lignes et colonnes (196x420). Deux méthodes différentes ont été utilisées pour réaliser les cartes des variables éco-géographiques :

- Pour tenir compte de la variabilité des variables non persistantes, la moyenne des valeurs mesurées a été calculée, dans une cellule donnée, à chaque échantillon de présence/absence de 10 minutes. Cette procédure a été répétée dans toutes les cellules contenant au moins un échantillon. Les valeurs à chaque point d'échantillonnage sont rasterisées par pondération par la variable (outils « rasterize »). On effectue une moyenne de ces valeurs par hexagone de la grille (outils « Statistique de zone »), cette valeur est enregistrée dans la table attributaire de la grille. On crée ensuite une couche centroïde de la grille contenant les moyennes des variables par maille. Les cartes biogéographiques finales pour les variables non persistantes ont été créées en interpolant les centroïdes de chaque maille à l'aide de l'interpolation pondérée par la distance inverse (IDW). L'IDW est une méthode d'interpolation spatiale qui suppose que les valeurs des points proches sont plus similaires que les valeurs des points plus éloignés (Li et Heap, 2008). Par conséquent, elle estime les valeurs à des endroits inconnus en donnant un poids plus important aux points échantillonnés plus proches (Li et Heap, 2008 ; Lu et Wong, 2008). L'IDW est une méthode de calcul peu exigeante qui a déjà été utilisée avec succès pour prévoir les variables environnementales (Li et Heap, 2008).
- La nature immuable des variables persistantes a permis d'utiliser les échantillons de 10 minutes pour créer des rasters sans avoir besoin d'utiliser les valeurs moyennes dans une grille. L'interpolation IDW a été utilisée pour générer les cartes biogéographiques des variables persistantes en interpolant directement toutes les valeurs des échantillons de 10 minutes. Pour finir, tous les rasters ont été transformés au format IDRISI (.rst).

## 5.2. Choix des types de modèles

### *L'analyse des facteurs de niche écologique ( Environmental Niche Factor Analysis - ENFA)*

Cette méthode compare la distribution observée de l'espèce étudiée dans l'hyper-espace formé par l'ensemble des variables éco-géographiques (VEG) sur l'intégralité de la zone d'étude (Hirzel *et al.*, 2002). Les habitats potentiels vont être construits en fonction de leur distance par rapport aux observations dans l'hyper-volume, cette technique partant alors du principe qu'une espèce a une distribution non aléatoire et que la majorité des individus va occuper globalement les mêmes amplitudes de valeurs de chaque VEG. A partir de points de présence seulement, le modèle ENFA calcule la pertinence de chaque zone basée sur la taille de la niche de l'espèce et va générer des cartes d'habitat potentiel (Davies *et al.*, 2008; Hirzel et Arlettaz, 2003; Hirzel *et al.*, 2002). La modélisation d'habitat avec la méthode ENFA est réalisée en utilisant le logiciel BIOMAPPER.4, mis à disposition gratuitement par l'Université de Lausanne (Dr. Alexandre Hirzel)<sup>7</sup>.

#### ➤ Analyse des données

Les cartes biogéographiques ont été normalisées à l'aide d'un algorithme de transformation Box-Cox (Hirzel *et al.*, 2002). Les rasters normalisés Pente, Distance à la Côte et Type de Substrat contenaient une trop grande plage de valeurs, ce qui entraînait des complications pour la suite de l'analyse. Les cartes originales ont donc été utilisées pour les étapes suivantes (Hirzel *et al.*, 2004).

---

<sup>7</sup> <http://www2.unil.ch/biomapper/products.html>

Afin de vérifier la colinéarité entre toutes les variables biogéographiques, une matrice de corrélation a été calculée et une première analyse des facteurs réalisée. Le tableau des valeurs propres de l'analyse des facteurs nous informe sur l'influence des variables sur la distribution : si aucune valeur propre n'est proche de 0, toutes les variables gardées expliquent en partie la distribution des dauphins. Un coefficient de corrélation de Pearson ( $|r|$ ) a été calculé et la valeur de 0,7 a été choisie comme seuil au-delà duquel la colinéarité des variables entraîne un manque de fiabilité du modèle (distorsion des estimations du modèle et des prédictions) (Dormann et al. 2013). Les variables ont donc été considérées comme fortement corrélées entre elles lorsque  $|r| > 0,7$ . Dans ce cas, une seule des variables corrélées est retenue pour éviter la redondance d'information. Les variables les plus pertinentes d'un point de vue écologique et celles ayant une plus grande influence sur la distribution des dauphins ont été conservées.

L'étape suivante consiste à réaliser une analyse factorielle qui réduit les VEG à quelques facteurs non corrélés en utilisant le calcul des vecteurs propres (Eigen system), qui sont les nouveaux axes (facteurs) de l'hyper-espace formé par les variables (Hirzel *et al.*, 2002). Les facteurs de l'analyse factorielle ENFA ont également des significations écologiques : le premier facteur explique ainsi la Marginalité (M) définie par Basille *et al.* (2008) comme "la différence entre les conditions utilisées en moyenne par l'espèce et les conditions disponibles dans la zone d'étude" et est calculée en mesurant la distance au carré du centroïde de l'habitat par rapport au centroïde de l'habitat disponible total. La marginalité varie entre 0 (l'espèce utilise des conditions similaires à celles des conditions moyennes disponibles) et 1 (l'espèce occupe un habitat spécifique dans la zone d'étude) (Hirzel *et al.*, 2002).

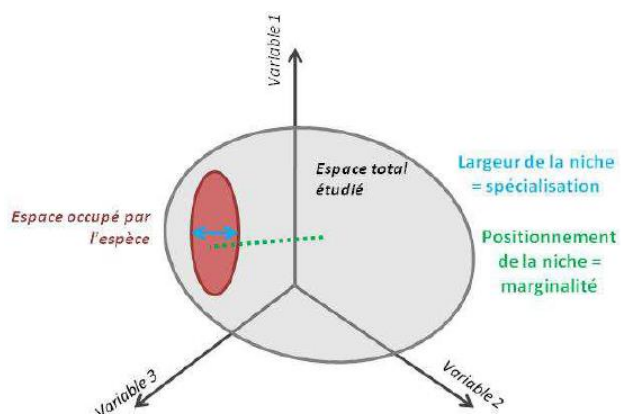


Figure 3. ENFA, marginalité et spécialisation, modifié d'après Hirzel (2002)

Les autres facteurs expliquent la Spécialisation (S) qui est la différence entre la variance de l'espèce et la variance globale. Elle peut être considérée comme la sensibilité de l'espèce aux variations des conditions environnementales (Figure 3). L'espèce occupant une niche étroite aura une spécialisation élevée (très sélective par rapport aux valeurs des VEGs) alors qu'une espèce occupant une niche plus large aura une spécialisation faible (Basille *et al.*, 2008). La spécialisation est difficile à interpréter, car elle varie de 0 à l'infini.

Cependant, une valeur supérieure à 1 indique un certain degré de spécialisation (Hirzel *et al.*, 2002).

Pour choisir le nombre de facteurs à retenir pour la suite, la part de l'information expliquée en fonction du nombre de facteurs retenus a été utilisée. Une fois le nombre de facteurs choisis, Biomapper crée un nombre de cartes temporaires (une par facteur), qu'il va ensuite compiler pour avoir la cartographie globale d'habitat potentiel. Différents algorithmes sont disponibles pour calculer la probabilité qu'à chaque pixel de présenter un habitat adéquat pour l'espèce étudiée: l'algorithme médian, l'algorithme de la distance géométrique moyenne, l'algorithme de la moyenne harmonique et l'algorithme de la distance minimale. Selon l'auteur, ce dernier ne fournit jamais de bons résultats, et l'algorithme de la moyenne harmonique a tendance à sur-ajuster les données. On comparera donc les deux premiers algorithmes (Annexe 10).

Les valeurs de marginalité et de spécialisation ont été utilisées pour calculer un indice d'habitat potentiel (HSI), qui a ensuite été utilisé pour créer la carte d'habitat potentiel (Hirzel *et al.*, 2006 ; Hirzel

et Arlettaz, 2003). L'HSI varie de 0 à 100, les valeurs inférieures signifiant une faible potentialité et les valeurs supérieures une potentialité d'habitat élevée.

Afin de choisir les meilleurs modèles parmi les différentes possibilités à tester et pour évaluer l'erreur de prédiction des modèles, la méthode cross validation "k-fold" a été utilisée. Cette méthode divise les données en plusieurs partitions de taille égale k, utilise k-1 partitions comme calibration du modèle et la partition restante pour le valider (Hastie *et al.*, 2001). Cette procédure est effectuée k fois, chacune d'entre elles utilisant un ensemble différent pour valider le modèle. Dans cette étude, une validation croisée (k=5) a été utilisée (Hirzel *et al.*, 2006)

Le pouvoir prédictif du modèle a été évalué avec l'indice de Boyce (Boyce *et al.*, 2002). La plage d'indices va de -1 à 1, des valeurs positives montrant la cohérence du modèle avec l'ensemble des données utilisées pour son évaluation. De même, des valeurs proches de 1 indiquent que la distribution calculée diffère d'une distribution attendue par hasard (Hirzel *et al.*, 2006).

À partir de l'indice de Boyce, une méthode basée sur le seuil a été développée pour évaluer la capacité du modèle à prédire la potentialité de l'habitat (Hirzel *et al.*, 2006). Pour ce faire, le HSI a été divisé en différentes classes, et pour chacune d'entre elles deux fréquences ont été calculées : (1) la fréquence prédite ( $P_i$ ), qui est le nombre de points d'évaluation prédits par le modèle dans chaque classe, divisé par le nombre total de points d'évaluation ; et (2) la fréquence attendue ( $E_i$ ), qui divise la superficie d'une classe d'adéquation de l'habitat par la superficie totale de l'étude (Hirzel *et al.*, 2006). Avec ces informations, le rapport prévu/attendu ( $P_i/E_i$ ) a été calculé pour chaque classe.

Pour évaluer le modèle, le rapport  $P_i/E_i$  a été calculé tout au long de la gamme HSI, générant 5 courbes  $P_i/E_i$  continues, une pour chacun des ensembles utilisés dans le processus de validation croisée. Trois aspects des courbes  $P_i/E_i$  ont été utilisés pour évaluer la précision du modèle : (1) la variance entre les courbes comme indication de la robustesse du modèle ; (2) la forme comme résolution des prédictions du modèle ; et (3) le maximum comme déviation du modèle par rapport à une attente aléatoire (Hirzel *et al.*, 2006).

Les courbes  $P_i/E_i$  ont été utilisées pour générer des seuils afin de diviser l'habitat en 4 classes différentes selon son adéquation :

- habitat inadapté : le rapport  $P_i/E_i$  est inférieur à 1,
- habitat marginal : le rapport  $P_i/E_i$  proche de 1,
- habitat adapté : le rapport  $P_i/E_i$  augmente de manière exponentielle,
- habitat optimal : le rapport  $P_i/E_i$  est le plus élevé

#### *Modèles Linéaires Généralisés (GLM) et Modèles Additifs Généralisés (GAM)*

Pour étudier la distribution des grands dauphins et les habitats préférentiels, deux algorithmes de modélisation différents ont été mis en œuvre dans le package Biomod2 dans R v.1.1.463 (Thuiller *et al.*, 2009). Ces algorithmes de modélisation sont deux méthodes de régression, des modèles linéaires généralisés (GLMs, McCullagh et Nelder, 1989) et des modèles additifs généralisés (GAMs, Guisan *et al.*, 2002). Ces algorithmes de modélisation fonctionnent bien avec les SDM (Guisan *et al.*, 2002; Phillips *et al.*, 2006; Franklin, 2010) et permettent une comparaison entre deux approches de modélisation différentes.

Les modèles linéaires généralisés (GLM) sont utiles pour ajuster les relations linéaires aux distributions de données non gaussiennes telles que les données de présence/absence (Hirzel *et al.*,

2001 ; McCullagh et Nelder, 1989). Les GLM relient la variable dépendante à une combinaison linéaire de variables explicatives (Hirzel *et al.*, 2001). Le principal avantage des GLM par rapport aux régressions linéaires classiques est la prise en compte partielle d'effets non linéaires à travers la fonction de lien, qui transforme la structure de dépendance initialement linéaire entre la variable réponse et les prédicteurs. Cette fonction est toutefois régie par des contraintes de régularité fortes qui restreignent le champ des effets modélisables. Elle présente notamment la spécificité d'être strictement monotone, ce qui implique que l'effet modélisé d'un prédicteur donné sur la variable de sortie est systématiquement de signe constant. C'est pour ces raisons théoriques qu'il paraît intéressant d'expérimenter, outre les GLM, les outils plus complexes offerts par les GAM. Les GAM (Hastie et Tibshirani, 1990) sont des extensions semi-paramétriques des GLM. Lorsque les données sont liées à certaines variables mais que les relations ne sont pas simplement linéaires, la modélisation additive peut être un outil utile pour améliorer la précision des prévisions. L'un des avantages de l'utilisation des GAM est qu'ils permettent des relations plus « souples » entre les variables de prédiction et de réponse (Hastie et Tibshirani, 1990). Les courbes GAM peuvent prendre des formes plus flexibles (asymétries ou ruptures de pente) et peuvent donc fournir un ajustement plus précis que les GLM (Hirzel, 2008).

Les GLM et le GAM ont été construits séparément, en utilisant une distribution d'erreur binomiale et la fonction de lien logit. En l'absence de données d'évaluation, une méthode de validation croisée ( $k=10$ , nombre contrôlé par l'argument "NbRunEval") a été effectuée. La fonction « DataSplit » du package nous permet de diviser notre ensemble de données de manière aléatoire en deux sous-ensembles. Une répartition des données de 75/25% pour l'étalonnage et le test des modèles respectifs a donc été mise en œuvre (Thuiller *et al.*, 2009). Construire des modèles avec toutes les informations disponibles peut être utile dans certains cas particuliers (par exemple, les espèces rares avec peu de points de présence) en utilisant la commande « do.full.model=T ». Le principal inconvénient de cette méthode est que, si on ne fournit pas de données séparées pour l'évaluation des modèles, les modèles seront évalués avec les mêmes données que celles utilisées pour la calibration. Cela peut conduire à des scores d'évaluation trop optimistes, c'est pourquoi cette commande ne sera pas utilisée dans cette étude. Une sélection de pseudo-absences a été effectuée avec la commande « NbRepPA ». Cet argument doit être corrélé avec l'utilisation de répétitions pour la calibration : une fois que le pool de pseudo-absences potentielles a été défini par la stratégie choisie, ici 75/25, un nombre défini par l'utilisateur (Nb.argument des absences) est choisi au hasard dans ce groupe. Nous avons donc un effet aléatoire dans le processus de calibrage provenant de la création de pseudo-absences pour nos données. L'argument NbRepPA définira un certain nombre de répétitions pour le retrait aléatoire des absences afin de constituer les ensembles de données de calibration. Il faut considérer que le nombre total de répétitions sera une multiplication des deux arguments de répétition. Ici, le nombre de répétitions sera de 2 (Barbet-Massin *et al.*, 2012).

L'importance des VEG a été calculée en utilisant une procédure de randomisation dans Biomod2 basée sur 10 runs de permutation (Thuiller *et al.*, 2009). Cette méthode est indépendante de la technique de modélisation et permet donc une comparaison directe entre les modèles. La procédure calcule la corrélation entre les prédictions standard (c'est-à-dire les valeurs ajustées) et les prédictions où une variable a été permutée de façon aléatoire. Une corrélation élevée (peu de différence entre 2 prédictions) indique que la variable n'est pas importante dans le modèle, et inversement, une faible corrélation indique que la variable est importante. Les variables sont classées de 0 à 1 selon le coefficient de corrélation moyen ; la variable la mieux classée a le plus d'influence sur le modèle, tandis qu'une valeur de 0 suppose que cette variable n'a aucune influence (Thuiller *et al.*, 2009). Les SDM basés sur des données de présence-absence peuvent entraîner des erreurs de commission et d'omission (Guisan et Thuiller, 2005 ; Franklin, 2010). Les erreurs de commission surviennent lorsque

Les modèles prédisent la présence d'une espèce dans des zones où elle n'est pas présente (faux positifs), tandis que les erreurs d'omission ne prédisent pas la présence d'une espèce dans des zones où elle est présente (faux négatifs) (Guisan et Thuiller, 2005). De telles erreurs peuvent être coûteuses, en particulier lorsqu'on essaie de délimiter l'habitat préférentiel d'une espèce ; par exemple, il pourrait être coûteux de protéger un habitat qui n'est pas important (Franklin, 2010).

Pour évaluer les performances et la précision des prévisions des SDM et comparer les méthodes de modélisation, la mesure de l'aire sous la courbe (AUC) de la courbe ROC (Receiver Operating Characteristic) (Annexe 11) a été utilisée dans Biomod2. L'AUC est une mesure indépendante du seuil du rapport entre les valeurs de présence/absence observées et les prédictions du modèle (Fielding et Bell, 1997). Les valeurs de AUC varient de 0 à 1 ; un score de 1 indiquant un ajustement parfait, des scores de 0,9 à 1 représentant un excellent ajustement, des scores de 0,8 à 0,9 représentant un bon ajustement, et ces scores de 0,5 à 0,6 indiquant que les prédictions du modèle ne sont pas meilleures que ce à quoi on pourrait s'attendre par hasard (Swets, 1988).

Les algorithmes de modélisation ont ensuite été combinés pour générer deux prédictions d'ensemble (une pour les GAM, une pour les GLM) de la présence de dauphins dans la zone d'étude (Thuiller *et al.*, 2009). Les GLM et les GAM individuels ont été pondérés en fonction de leur précision prédictive (plus la note d'évaluation est élevée, plus le modèle a de poids) grâce à l'argument « prob.mean.weight.decay » (Marmion *et al.*, 2009) ; seuls les modèles ayant des valeurs d'AUC supérieures à 0,8 ont été utilisés pour créer les modèles d'ensemble. Les modèles d'ensemble ont ensuite été utilisés pour fournir une sortie visuelle de la probabilité d'occurrence des espèces, où les valeurs vont de 0 à 1, indiquant respectivement la plus faible et la plus forte probabilité d'occurrence. Enfin, nous avons utilisé les valeurs AUC pour tester si les modèles d'ensemble finaux étaient plus performants que leurs modèles individuels correspondants (Marmion *et al.*, 2009).

### III. Résultats

#### 1. Reformatage des données et incrémentation de la base de données existante

##### ➤ Formatage données GPS

Un gros travail d'homogénéisation des données a été effectué pour les données GPS. En effet, les données des années 2016 et 2017 étaient lisibles sur le SIG mais les coordonnées géographiques n'étaient pas visibles dans les tables attributaires, empêchant l'exploitation des données. Pour ces deux années, chaque trace GPS au format gpx a été implémentée et modifiée une à une dans QGIS pour obtenir un format shapefile et des tables attributaires modifiables et comparables. Une fois homogénéisées, les couches de tracés shapefiles ont été fusionnées par année et transformées en fichier csv. Les sorties en doublons (deux bateaux pour la même observation) et les tracés disposant de trop peu de points ou de points trop éparpillés (dysfonctionnement du GPS) ont été retirés pour ne pas fausser le calcul de l'effort de prospection.

##### ➤ Enrichissement et valorisation de la base de données 2019

Concernant l'export des données de la plateforme ObsenMer (2019), une fonctionnalité permet d'exporter séparément le fichier contenant les observations et le fichier contenant les trackings gps des sorties. La mise en forme de ces fichiers s'est déroulé sur Excel. Les deux jeux de données ont été triés, puis chaque sortie renommée dans chaque fichier avec le même numéro identifiant.

Pour les données de 2016, chaque sortie a aussi été renommée de la même manière sur le fichier .csv de tracking issus de la transformation des tracés gpx ainsi que sur le fichier contenant les observations. Toutes ces données ainsi que les données manquantes récupérées en amont ont ensuite été implémentées dans la base de données de l'association. Une attention particulière a été portée aux changements de format entre années (UTC, unité de mesure ect.) afin d'homogénéiser les informations.

➤ Transformation des données pour avoir de la présence/absence

Une colonne « Présence » a été créée sur le fichier de tracking, contenant un 0 pour une absence et un 1 pour chaque point gps où une observation ponctuelle de dauphin a été enregistrée sur le fichier « Observation ». Pour les observations de type suivi, tous les points de trackings entre les évènements « Début de suivi » et « Fin de suivi » ont été codé d'un 1. Ce travail de recodage a été effectué manuellement en recoupant les informations des deux fichiers ObsenMer.

Pour homogénéiser les points d'échantillonnage au niveau temporel, le fichier contenant les points de tracking a été découpé de manière à obtenir des points GPS avec des intervalles de temps égaux de 10 min (correspondant à la traversée à 12 nœuds d'une maille de la grille sur SIG). Cela permet d'avoir des points régulièrement répartis sur la ligne de transect et de ne pas avoir un point d'absence et un point de présence situés dans une même cellule du raster. Ce jeu de données est ensuite rééchantillonné en sélectionnant aléatoirement 735 points d'absences sur les 7183 disponibles pour avoir un ratio 1 :1 avec les 735 présences disponibles (package « dplyr », commande « sample »). Après traitement, le jeu de données est composé de 735 points d'absences et 735 points de présence. Un autre jeu de données d'absence sera généré lors de la modélisation en sélectionnant aléatoirement 735 pseudo-absences (voir p.16 ).

## 2. Résultats des analyses préliminaires

➤ Effort de prospection

De mars 2016 à novembre 2019, 741 sorties ont été réalisées. La distance totale parcourue est de 16247,2 km. Ses 741 sorties représentent un total de 494 observations de Grands Dauphins. Au cours des quatre années d'étude, l'effort a été réparti dans le secteur d'étude comme montré sur les figures 4 &5. L'ensemble de la zone d'étude a été couvert de manière homogène entre 2016 et 2019, avec un gradient correspondant à la distance au port. Ainsi, le secteur autour de la pointe du Grouin a été prospecté de manière plus marquée. La

Figure 5 présente la répartition de l'effort de prospection en kilomètres par an. Le test ANOVA montre que les variations annuelles (ANOVA :  $F = 34,176$ ,  $P = 0.02804$ ) d'effort de prospection sont significatives.

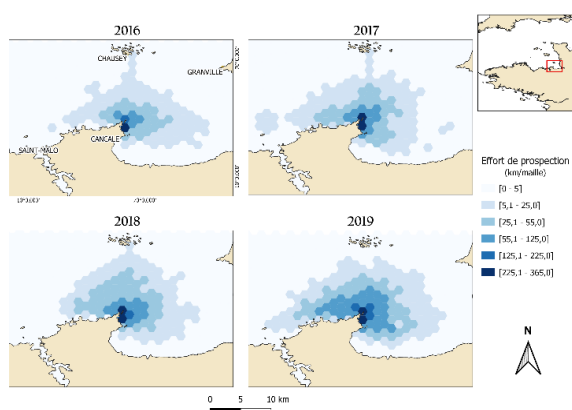


Figure 4. Carte représentant l'effort de prospection, en nombre de km par maille hexagonale, pour les années 2016,2017,2018,2019

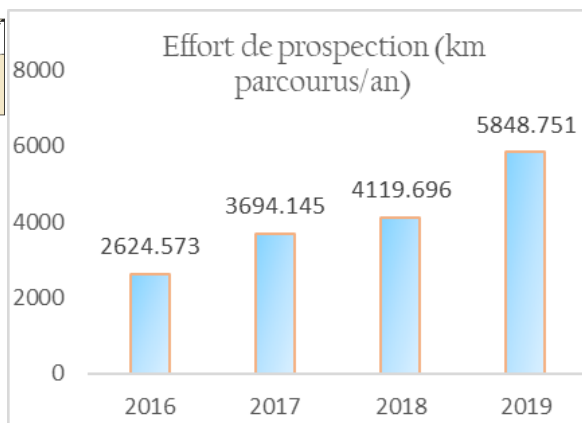


Figure 5. Effort de prospection, en nombre de km parcourus par maille, de 2016 à 2019

➤ Répartition spatiale et temporelle

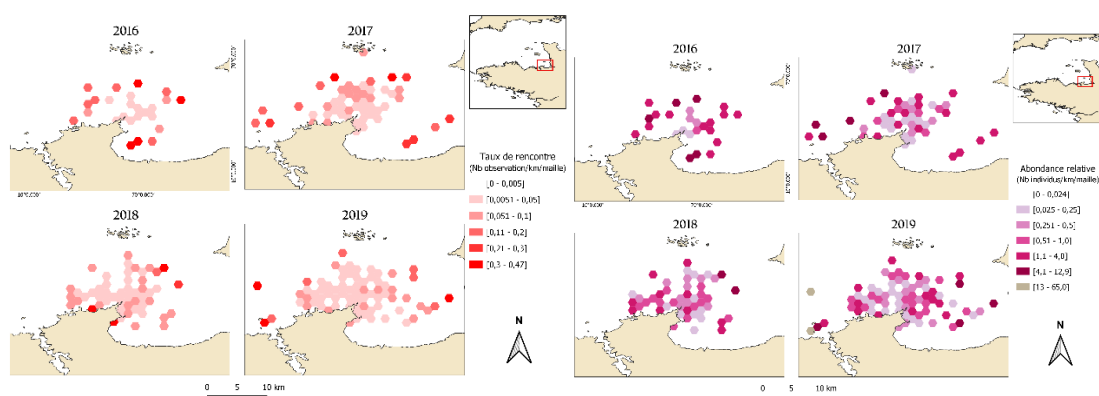


Figure 6. a) Carte représentant le taux de rencontre, en nombre d'observation par km par maille hexagonale, pour les années 2016, 2017, 2018 et 2019 b) Carte représentant l'abondance relative, en nombre d'individus par km par maille hexagonale, pour les années 2016, 2017, 2018 et 2019

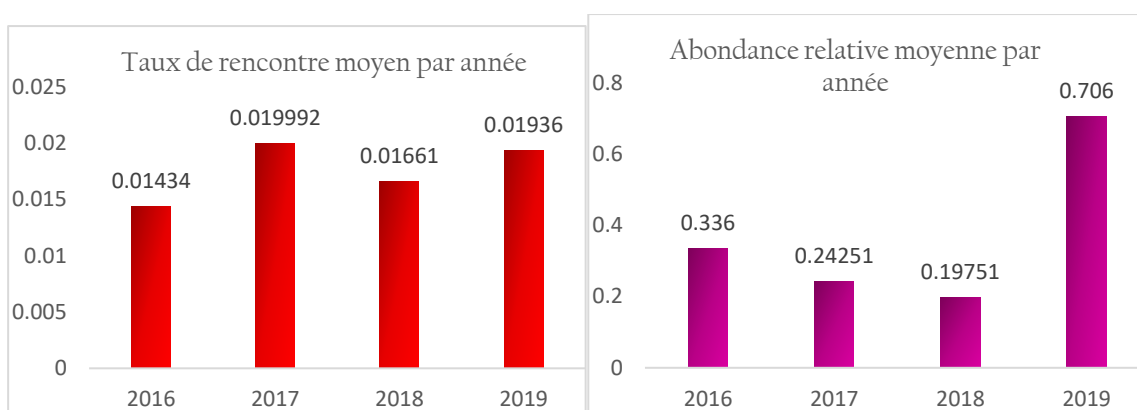


Figure 7. Taux de rencontre moyen et abondance relative moyenne par année

On peut noter que le taux de rencontre (TR) était un peu plus élevé en 2017 que les autres années. Au cours de la période d'étude, tout mois confondus, le taux de rencontre des grands dauphins était relativement fort avec un maximum de 47 observations par 100 km d'effort en 2017 et 2019 dans les extrémités de la zone d'étude (Figure 6 a). Il faut cependant bien noter qu'une à deux observations maximums ont été effectuées dans ces mailles, pour un effort faible. Hormis ces cas extrêmes, le taux de rencontre reste homogène entre 2017 et 2018 sur la zone d'étude (Figure 6), avec une moyenne de  $2 \pm 0,003$  groupes par 100 km d'effort (IC 95%). Néanmoins, les tests ANOVA démontrent que les variations annuelles (ANOVA :  $F = 1,0023$ ,  $P = 0.4222$ ) des taux de rencontre (Figure 7) ne sont pas significatives.

Au cours de la période d'étude, toutes années confondues, l'abondance relative (AR) des grands dauphins a varié de 0 à 13 individus par kilomètre d'effort. Les mailles avec les plus grandes valeurs d'abondance relative sont situées en face de Saint-Malo et vers les îles Chausey, zones peu prospectées mais où de grands groupes ont été observés. Hormis ces zones, l'AR reste relativement homogène sur toute la zone d'étude (Moyenne :  $0,37 \pm 0,23$  individus / km) (Figure 6). L'abondance relative était plus élevée en 2019 par rapport aux autres années (Figure 7) néanmoins les variations annuelles (ANOVA :  $F = 1,0975$ ,  $P = 0.4048$ ) des abondances relatives ne sont pas significatives.

### 3. Modélisation

#### ➤ Analyse préliminaire des données environnementales

La matrice de corrélation de Spearman (Figure 8 + Annexe 12) et l'ACP (Figure 9) ont permis de définir l'importance des variables environnementales et de montrer celles qui sont fortement corrélées, c'est-à-dire la bathymétrie avec la profondeur, la salinité négativement avec la chlorophylle ainsi que les variables calculées à partir des données bathymétriques du GEBCO (TRI, Pente, Rugosité). La distance à la côte est aussi fortement corrélée à ces variables. On choisira donc une seule variable par groupe de corrélation pour la suite des analyses en fonction de leur importance dans les différents modèles utilisés.

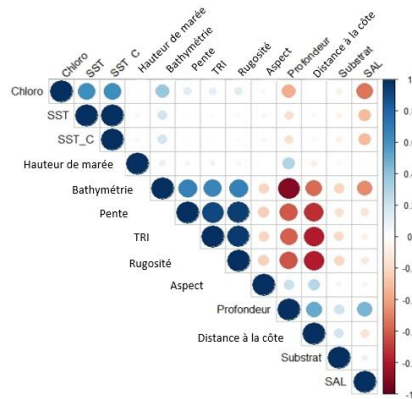


Figure 8. Matrice de corrélation de Spearman

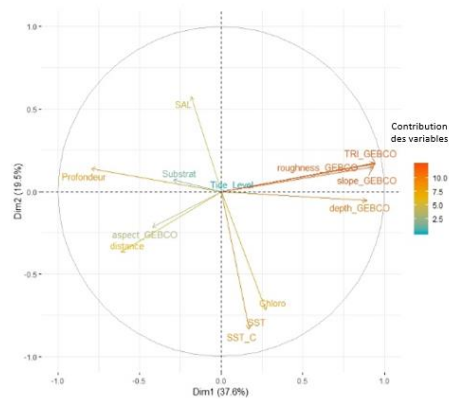


Figure 9. ACP

#### ➤ Transformation des données environnementales

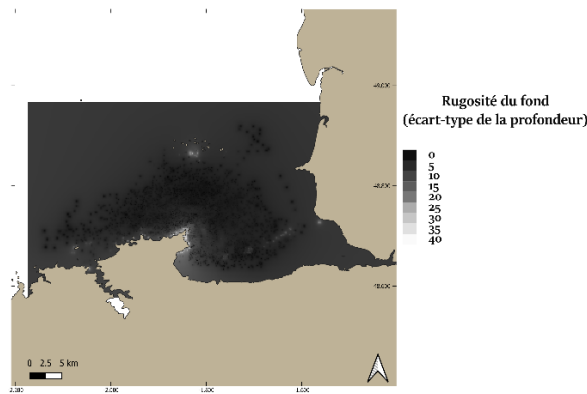


Figure 10. Carte éco-géographique de la rugosité du fond

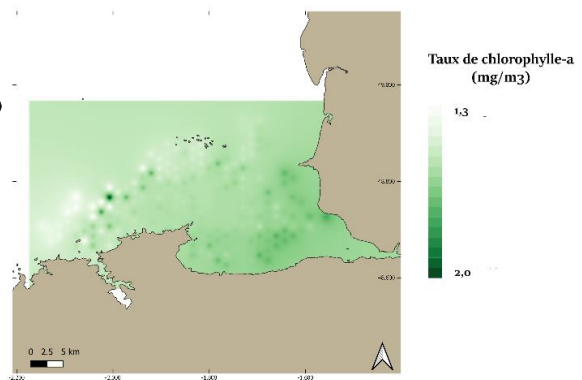
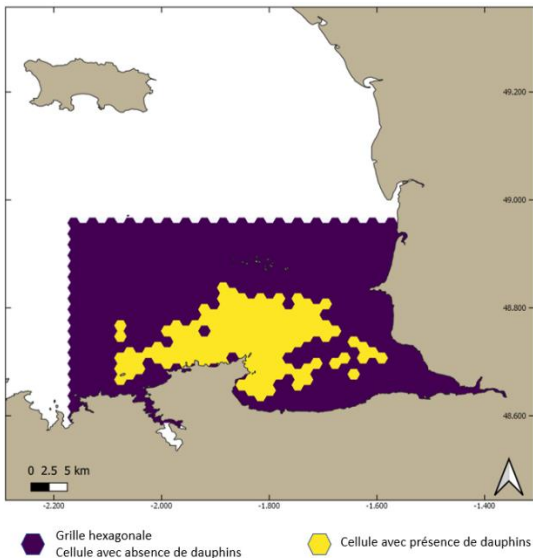


Figure 11. Carte éco-géographique de la concentration en chlorophylle-a

Les 12 variables éco-géographiques ont été transformées en fonction de leur caractère persistant ou non persistant en suivant le protocole développé plus haut (p.13) afin d'être implantés dans le logiciel Biomapper4 ( Figure 10 & 11 + Annexe 13 )





## Raster booléen

Figure 12. Carte montrant la grille hexagonale utilisée pour créer la carte d'espèce (raster booléen) et les carte éco-géographiques pour les variables persistantes et non persistantes.

La carte (Figure 12) est un raster booléen créé comme décrit précédemment (p.13) Les cellules de la maille où une observation de dauphin a été échantillonné sont de valeur 1, celles où aucune n'a été échantillonné de valeur 0.

## Résultats de l'ENFA

### o Choix des variables

Dans l'analyse préliminaire des données, une matrice de colinéarité a été générée pour étudier la colinéarité entre les variables éco-géographiques. La matrice montre que 5 variables sont fortement corrélées. Les variables identifiées par le modèle comme ayant la plus grande influence sur la distribution de l'espèce sont gardées et celles du même groupe ayant moins d'influence sont retirées du modèle final. De ce fait, la salinité, la bathymétrie, la pente et la valeur TRI de rugosité ont été retirées de l'analyse. La chlorophylle-a et la température ont été gardées car elles ont une plus grande influence sur la distribution de l'espèce que la salinité (

Tableau 2), et parce que ces deux variables apportent davantage d'informations sur la distribution, selon les connaissances sur l'espèce cible (Smith et al., 2008; Davis et al., 2002).

	Chloro-box	Profondeur-box	Roughness-box	SST°C-box	Depth-box	TideLevel-box	TRI-box	Distance	Salinity-box	Slope	Substrat	Aspect-box
Chloro-box	1.000											
Profondeur-box	<b>-0.779</b>	1.000										
Roughness-box	-0.014	-0.219	1.000									
SST°C-box	<b>0.772</b>	<b>-0.714</b>	0.054	1.000								
Depth-box	0.515	<b>-0.785</b>	0.579	0.58	1.000							
Tide-Level-box	0.414	-0.463	0.008	0.425	0.448	1.000						
TRI-box	-0.027	-0.207	<b>0.991</b>	0.042	0.556	0.001	1.000					
Distance	0.249	-0.005	-0.492	0.225	-0.159	0.222	-0.479	1.000				
Salinity-box	<b>-0.891</b>	<b>0.812</b>	-0.009	<b>-0.828</b>	-0.591	-0.474	0.001	-0.332	1.000			
Slope	-0.009	-0.188	<b>0.954</b>	0.059	0.59	0.018	<b>0.934</b>	-0.47	-0.002	1.000		
Substrat	-0.216	0.297	-0.081	-0.229	-0.311	-0.304	-0.078	-0.175	0.278	-0.07	1.000	
Aspect-box	0.321	-0.287	-0.096	0.2	0.163	0.328	-0.103	0.369	-0.331	-0.05	-0.18	1.000

Tableau 2. Matrice de corrélation montrant la colinéarité entre les VEG. Deux variables sont considérés hautement corrélés quand  $r > 0.7$  (en gras).

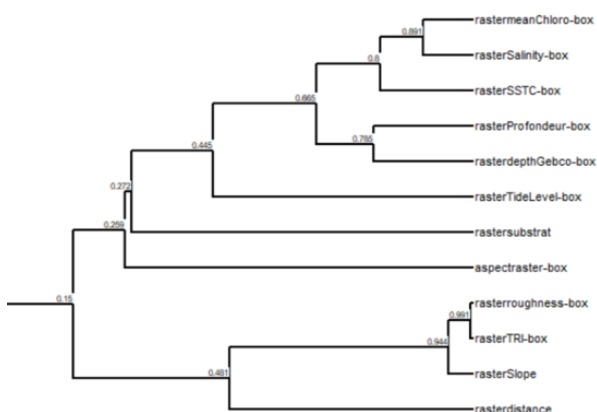


Figure 13. Arbre des corrélations

	Valeur propre	Expl.Spec.	Cum.Expl. Spécialisation
Chlorophylle_a	0.674	0.081	0.081
Profondeur	2.502	0.301	0.382
Rugosité	1.658	0.199	0.581
SST °C	0.733	0.088	0.67
Bathymétrie	0.62	0.075	0.744
Hauteur de marée	0.613	0.074	0.818
TRI	0.351	0.042	0.86
Distance à la côte	0.317	0.038	0.898
Salinité	0.234	0.028	0.926
Pente	0.22	0.026	0.953
Substrat	0.203	0.024	0.977
Aspect	0.189	0.023	1.000

Tableau 3. Tableau des valeurs propres ENFA toutes variables

Deux ENFA comprenant un nombre différent de variables ont été réalisées et comparées :

- ENFA à 8 variables : chlorophylle, Profondeur, Rugosité, Température, Tide level (Hauteur de marée), Distance à la côte, Type de substrat et Aspect.
- ENFA à 6 variables : idem sans Aspect et Substrat.
  - Analyse factorielle

Une fois les ENFA réalisées, un tableau de résultats (Tableau 4) montre, pour chacun des facteurs, quelles variables agissent le plus sur la distribution des dauphins, soit sur leur marginalité (facteur 1), soit sur leur spécialisation (facteur 2 et plus), en présentant un score plus élevé. Le signe des variables pour la marginalité indique si celles-ci sont corrélées positivement ou négativement avec la distribution des dauphins. En revanche, le signe obtenu pour les axes de spécialisation n'apporte aucune information particulière.

▪ ENFA 8 variables

VEG	Marginality (11%)	SF1 (33%)	SF2 (19%)	SF3 (12%)	SF4 (11%)	SF5 (7%)	SF6 (4%)	SF7(3%)
Chlorophylle	0.005	0.849	0.258	0.555	-0.549	-0.281	-0.042	-0.315
Profondeur	0.3	-0.022	-0.459	0.052	-0.737	-0.232	-0.511	-0.132
Rugosité	-0.874	0.077	-0.146	0.186	-0.205	-0.148	-0.003	0.136
SST°C	-0.148	0.083	-0.642	-0.769	-0.031	-0.201	-0.263	0.189
Hauteur de marée	-0.193	-0.491	0.509	-0.207	-0.318	0.119	0.054	0.03
Distance	-0.039	0.159	-0.174	0.129	0.003	0.879	0.336	0.117
Substrat	0.293	-0.021	0.019	-0.013	-0.069	-0.108	0.44	0.696
Aspect	-0.01	0.01	0.005	0.063	0.088	-0.083	-0.599	0.574

Tableau 4. Résultats du modèle ENFA à 8 variables, montrant les scores de chaque VEG, la marginalité et les facteurs de spécialisation (SF) et le pourcentage d'information expliquée par chaque facteur

Les 6 premiers facteurs expliquent 92 % de la spécialisation totale (total sum of eigen values = 0.919). Le premier facteur explique la marginalité et 11% de la spécialisation. Le score de chaque VEG pour ce facteur révèle que la marginalité est particulièrement influencée par la profondeur et la rugosité, en montrant que la présence des grands dauphins est liée à une valeur haute de profondeur et une valeur faible de rugosité. Les autres VEG ont peu d'effet sur la marginalité. Les autres facteurs (SF) expliquent le reste de la spécialisation. Le premier facteur de spécialisation (SF1), qui compte pour 33% de la spécialisation, montre que le choix d'habitat des grands dauphins est surtout influencé par le taux de chlorophylle a (0.849), et la hauteur de la marée (0.491). Les autres facteurs montrent une certaine sensibilité à la profondeur (SF2, SF4), à la température (SF2, SF3), à la chlorophylle (SF3, SF4), et à la distance à la côte (SF5). La VEG avec la plus forte influence sur la spécialisation, en combinant les 5 premiers facteurs, est la chlorophylle (2.497). De plus, la marginalité et la spécialisation calculées par le modèle ENFA sont respectivement de 0.657 et 0.818, ce qui montre que les conditions d'habitat des grands dauphins est assez similaire aux conditions moyennes de la zone d'étude, et que l'espèce peut facilement s'adapter à des conditions environnementales variables.

▪ ENFA 6 variables

VEG	Marginality (12%)	SF1 (36%)	SF2 (21%)	SF3 (13%)	SF4 (11%)	SF5 (7%)
Chlorophylle	0.005	-0.855	-0.252	0.615	0.469	-0.332
Profondeur	0.314	0.02	0.462	0.159	0.783	-0.252
Roughness	-0.914	-0.082	0.153	0.203	0.176	-0.128
SST°C	-0.155	-0.083	0.643	-0.721	0.157	-0.158
Hauteur marée	-0.202	0.477	-0.506	-0.131	0.327	0.119
Distance	-0.041	-0.164	0.176	0.136	-0.07	0.878

Tableau 5. Résultats du modèle ENFA à 6 variables, montrant les scores de chaque VEG, la marginalité et les facteurs de spécialisation (SF) et le pourcentage d'information expliquée par chaque facteur

L'ENFA à 6 variables (Tableau 5) montre des résultats similaires à la précédente, avec des scores est légèrement plus élevés pour chaque VEG.

- Comparaison du nombre de facteurs à prendre en compte

Pour chacune des deux ENFA, on comparera les résultats obtenus par les modèles à 4 ou 5 facteurs.

Pour l'ENFA à 8 variables, le modèle à 4 facteurs explique 88 % de l'information totale (0.873) et 75 % de la spécialisation (0.746). Le modèle à 5 facteurs explique 92% (0.926) de l'information totale et 85% (0.853) de la spécialisation.

Concernant l'ENFA à 6 variables, le modèle à 4 facteurs explique 91% de l'information totale (0.909) et 82 % (0.818) de la spécialisation. Le modèle à 5 facteurs explique 96% (0.964) de l'information totale et 93% (0.927) de la spécialisation.

- Choix du modèle et du nombre de facteurs

A la suite de Hirzel *et al.* (2006), l'algorithmme de la moyenne géométrique a été choisi pour générer la carte d'adéquation des habitats, car il ne fait aucune hypothèse sur la répartition des espèces. Cette méthode tient compte de la proximité des points d'espèces dans l'espace environnemental et donne une meilleure adéquation là où les points d'espèces présentent une densité plus élevée (Hirzel et Arlettaz, 2003). De plus, après plusieurs tests de modèles, il s'avère que l'algorithmme de moyenne géométrique fournit toujours un meilleur indice de Boyce que l'algorithmme médian (Tableau 6).

	Indice de Boyce	
	Median	Geometric mean
ENFA 8 variables- 4 facteurs	0.575+- 0.1457	0.824 +- 0.108

Tableau 6. Tableau de comparaison entre les indices de Boyce pour les modèles issus des algorithmmes médian, et de moyenne géométrique

	Indice de BOYCE	
	Modèle à 4 facteurs	Modèle à 5 facteurs
ENFA 8 variables	0.824 +- 0.108	0.683 +- 0.40
ENFA 6 variables	<b>0.87 +- 0.1411</b>	0.83 +- 0.13

Tableau 7. Tableau de comparaison entre les indices de Boyce pour les modèles à 4 ou 5 facteurs

On retiendra le modèle d'ENFA à 6 variables, et 4 facteurs.

- Interprétation des courbes P/E

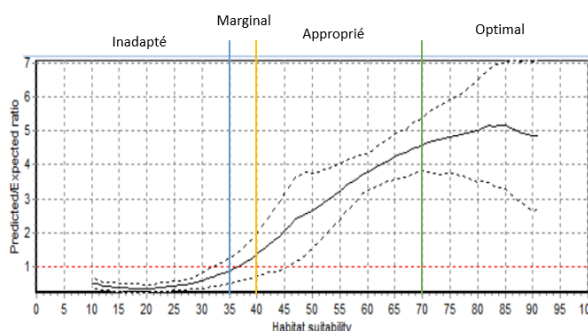


Figure 14. Graphique montrant la moyenne des courbes P/E pour le modèle et les quatre classes de probabilité d'habitat pour lesquelles l'habitat a été catégorisée

La cross validation (indice de Boyce = 0.87 +- 0.1411) (Tableau 7) et les courbes P/E (Annexe 15. Courbes P/E à 6variables 4 facteurs) ont montré que le modèle avait un bon pouvoir prédictif. La variance le long des courbes P/E a montré une augmentation constante avec l'indice HSI, avec un intervalle de confiance plus étroit pour les indices HSI plus faibles, indiquant que la puissance prédictive était plus précise pour les zones à faible adéquation (HSI < 35). Sur la base des courbes P/E(Figure 14), l'habitat a été classé en 4 classes différentes :

- (1) "habitat inadapté" pour les valeurs HSI inférieures à 35 ;
- (2) "habitat marginal" pour les valeurs HSI entre 35 et 40 ;
- (3) "habitat approprié" pour les valeurs HSI entre 40 et 70 ;
- (4) "habitat optimal" pour les valeurs HSI supérieures à 70.

La carte d'adéquation de l'habitat qui en résulte pour la zone d'étude est présentée ci-dessous (Figure 15).

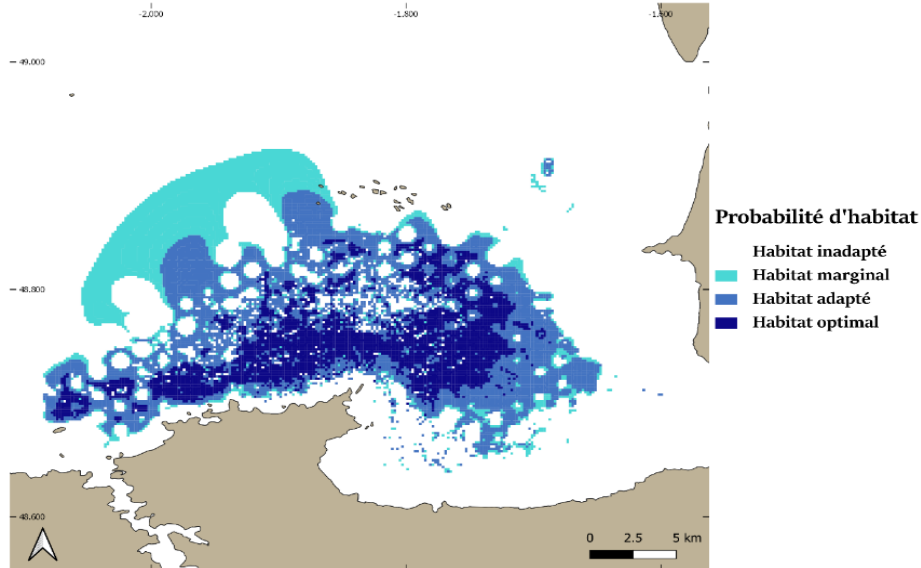


Figure 15. Carte de probabilité d'habitat des grands dauphins basée sur l'indice HSI calculé avec l'ENFA

La carte révèle que les eaux de fond de baie et les eaux côtières peu profondes de la zone d'étude ne conviennent pas à l'espèce. L'habitat marginal est lié à de petits endroits spécifiques autour des habitats adaptés, sauf le croissant à l'Est de l'archipel de Chausey, zone peu prospectée. Inversement, les eaux un peu plus profondes face à Saint Malo et entre la Pointe du Grouin et Chausey comprennent les habitats les plus appropriés pour les dauphins. Bien que certaines zones à l'intérieur de la baie comprennent des habitats adaptés et même optimaux pour l'espèce, la plupart des zones où les valeurs de rugosité de fond sont fortes ont été classées comme inadaptées.

### Résultat des GLM /GAM

#### ➤ Choix entre absence réelle et pseudo-absence

Type de modèle	GAM		GLM	
	Absences réelles	Moyenne des 2 runs de pseudo-absences	Absences réelles	Moyenne des 2 runs de pseudo-absences
6 variables	0.675 ± 0.02	0.964 ± 0.008	0.632 ± 0.02	0.931 ± 0.10
4 variables	0.657 ± 0.02	0.961 ± 0.007	0.623 ± 0.02	0.942 ± 0.01

Tableau 8. Comparaison des AUC entre les modèles utilisant de vraies absences et ceux utilisant des pseudos-absences

Afin de déterminer quelle méthode est la plus efficace entre utiliser les absences réelles échantillonnées sur le terrain et les pseudo-absences générées automatiquement par BIOMOD2, une comparaison de modèles a été effectuée (Annexe 16. Tableau récapitulatif des valeurs d'AUC pour les modèles GLM et GAM à 6 ou 4 variables selon la méthode de génération d'absence utilisée (absence réelle vs. Pseudos absences) (2 répétitions random pour la création des pseudos absences)). Les techniques de modélisation utilisant les pseudo-absences ont démontré un pouvoir prédictif bien supérieur aux modèles utilisant les absences réelles (AUC = 0.675 ± 0.02 en moyenne pour les modèles GAM 6 à variables utilisant les absences réelles contre 0.964 ± 0.008 en moyenne pour les modèles utilisant les pseudo-absences) (Tableau 8). Pour la suite des analyses, les modèles seront donc effectués en utilisant des données d'absences générées de manière aléatoire par le package (nommées ci-après « Pseudo-absences »).

➤ Choix des variables

Une comparaison de modèles (Annexe 17, Annexe 18) utilisant respectivement 4 ou 6 variables a été effectuée. Les résultats sont présentés ici comme les moyennes des valeurs d'importances des variables, pour les 20 modèles générés par méthode de modélisation (10 cross-validations x 2 répétitions aléatoires de sélection des pseudo-absences). Que ce soit pour la méthode GLM ou la méthode GAM, il apparaît que les variables TL et SST ont une faible influence sur les modèles (Tableau 9). Les retirer permettra d'obtenir un modèle plus simple permettant de se focaliser sur les variables environnementales les plus influentes.

Var. environnementales	GLM - 735 Présence, 735 Pseudo-absences		Var. environnementales	GAM - 735 Présence, 735 Pseudo-absences	
	Modèle 6 variables	Modèle 4 variables		Modèle 6 variables	Modèle 4 variables
Chlorophylle-a	0.05002	0.1862	Chlorophylle-a	0.11995	0.3512
Distance à la côte	<b>0.49014</b>	<b>0.4607</b>	Distance à la côte	<b>0.5127</b>	<b>0.46965</b>
Profondeur	0.2117	0.19975	Profondeur	0.14975	0.26205
Rugosité	<b>0.66765</b>	<b>0.581</b>	Rugosité	<b>0.587505</b>	<b>0.61485</b>
Hauteur marée	0.0176		Hauteur marée	0.0429	
SST (°C)	0.07805		SST (°C)	0.04945	

Tableau 9. Valeurs d'importance des variables pour les modèles à 6 ou 4 variables a) GLM b) GAM

➤ Comparaison entre modèles individuels et consensus

Les modèles utilisés pour l'analyse finale sont au nombre de 42 : 20 modèles individuels GAM et 20 modèles individuels GLM (10 cross-validations x 2 répétitions aléatoires) et deux modèles d'ensemble regroupant respectivement les 20 modèles GLM et les 20 modèles GAM. La figure (Figure 16) représente les valeurs d'AUC pour chaque modèle individuel et pour les modèles d'ensemble. Les valeurs d'AUC pour les modèles individuels indiquent que tous les modèles ont de bons pouvoirs prédictifs (>0.9). Les modèles GAM tendent à avoir un meilleur pouvoir prédictif que les GLM (moyenne GAM= 0.96, moyenne GLM=0.95). On remarque que les modèles d'ensemble ne sont pas forcément meilleurs que les modèles individuels. En effet, le meilleur modèle GLM a une valeur d'AUC de 0,966 contre 0,951 pour le modèle d'ensemble. Concernant les GAM, le meilleur modèle à un AUC de 0,980 contre 0,968 pour le modèle d'ensemble (Annexe 20).

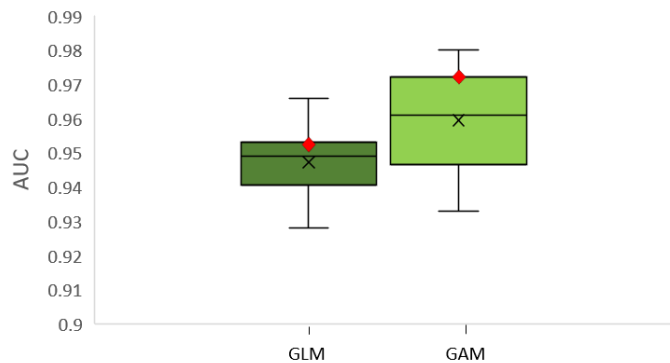


Figure 16. Différences de précision des modèles (AUC de la courbe ROC) entre les 10 cross validations runs et les 2 répétitions de pseudo-absences, pour chaque modèle individuel de distribution de l'espèce (boxplot) et chaque modèle d'ensemble (losange rouge). Les moustaches indiquent les plus hautes et faibles valeurs d'AUC des modèles individuels.

➤ Courbes réponses et importances des variables

La variable environnementale la plus influente pour les GAM et les GLM est la rugosité de fond (Tableau 10) Les courbes réponses issues des GAM et des GLM indiquent que les valeurs de rugosité inférieures à 5 et supérieures à 30 sont importantes pour prédire la présence des grands dauphins.

La distance à la côte semble aussi influencer de manière non négligeable la distribution des dauphins (Tableau 10) indiquant une préférence pour une proximité à la côte de moins de 4.3 km concernant les GAM (Figure 22), et de moins de 2.9 km concernant les GLM (Figure 21). La concentration en chlorophylle-a et la profondeur démontrent une

faible influence sur la prédiction d'occurrence de l'espèce. Néanmoins les courbes réponses indiquent une préférence pour les habitats avec une concentration chlorophyllienne inférieure à 1.5 mg/m<sup>3</sup> et une affinité pour les profondeurs supérieures à 12 m pour les GAM (Figure 22), et de 15m pour les GLM (Figure 21).

Variables environnementales	GLMs individuels	GAMs individuels
Chlorophylle-a	0.144	0.353
Distance à la côte	<b>0.483</b>	<b>0.441</b>
Profondeur	0.157	0.265
Rugosité	<b>0.63</b>	<b>0.624</b>

Tableau 10. Importance des VEG sur la distribution des grands dauphins du golfe normand-breton, selon les modélisations GAM et GLM. Les valeurs sont les moyennes des 20 modèles (Annexe 19) réalisés pour chaque technique de modélisation. En gras : variables avec le plus d'influence sur la distribution des grands dauphins.

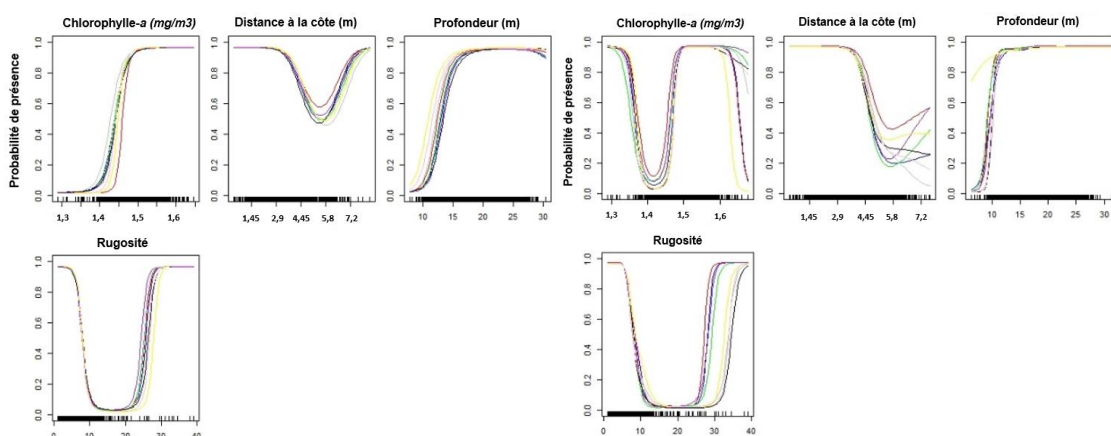


Figure 17. Courbes réponses pour les GLM

Figure 18. Courbes réponses pour les GAM

### ➤ Cartes de distribution

Utilisant ces données, les modèles d'ensemble (Figure 19) et les modèles individuels (Annexe 21 & 22) prédisent la présence de dauphins de façon prédominante à l'extrémité Nord-Est de la baie du Mont Saint Michel, dans des eaux plus profondes et plus près des côtes qu'à l'intérieur de la baie. Une forte probabilité de présence des dauphins est démontrée entre la Pointe du Grouin et Chausey, dans des zones où les valeurs de rugosité de fond et de concentrations en chlorophylle-a sont plus faibles. Le modèle d'ensemble GLM prédit la présence de dauphins au large de l'embouchure de la Rance, où de faibles valeurs de chlorophylle-a et de rugosité ont été enregistrées (Figure 10, Figure 11).

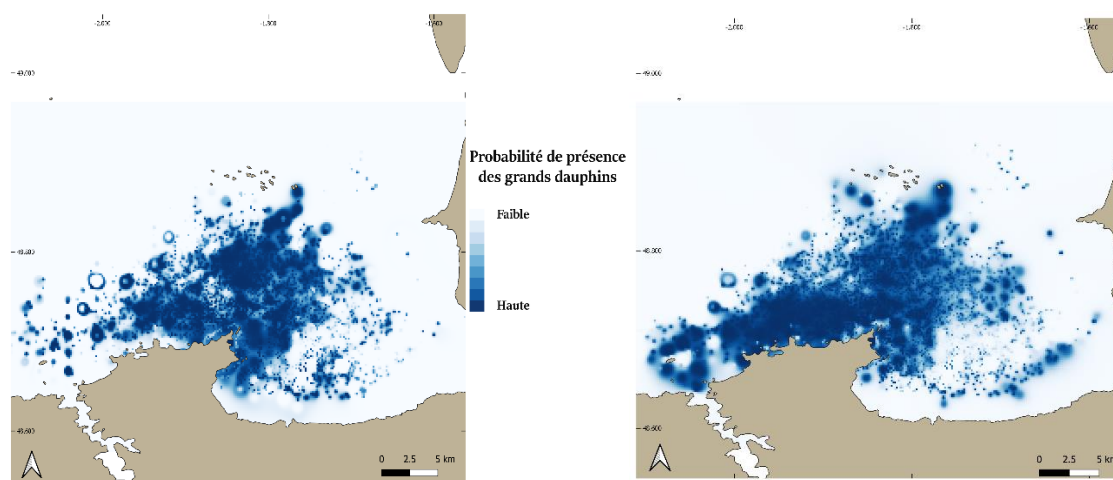


Figure 19. Cartes de probabilité de présence a) Modèle GAM d'ensemble b) Modèle GLM d'ensemble

## IV. Discussion

### ✓ Comparaison des variables environnementales influentes selon les différents modèles

Que ce soit pour l'ENFA, les GLM ou les GAM, les mêmes variables influentes ont été mises en lumière. En effet, la rugosité de fond est une des variables les plus influentes sur la distribution des grands dauphins, indiquant une plus grande probabilité de présence pour de faibles valeurs de rugosité. Ces résultats concordent avec les travaux de (Rodriguez-Ferrer *et al.*, 2020), qui démontrent que les grands dauphins ont été davantage observés voyageant dans des zones de faible rugosité. Le modèle semble tout de même suggérer qu'au-delà d'une certaine valeur de rugosité, la probabilité de présence augmente de nouveau. Cette information indique que les grands dauphins pourraient utiliser des zones de rugosité plus élevées. Ce résultat pourrait être lié à un comportement de recherche alimentaire, en effet des études en Irlande et en Ecosse ont démontré que les Grands Dauphins présentaient un comportement accru de recherche alimentaire dans les zones où la rugosité est importante (Hastie *et al.*, 2004).

Les résultats de l'ENFA ont indiqué que la profondeur de l'eau était le second prédicteur le plus important de la distribution des dauphins de la zone d'étude, et montrent que les eaux plus profondes de la baie constitueraient un habitat optimal pour eux. Il faut cependant noter que ces zones profondes correspondent aux zones de faible profondeur pour d'autres sites d'étude. Des études menées dans le sanctuaire Pelagos en Méditerranée (Gnone *et al.*, 2011; Cañadas *et al.*, 2002) et en Caroline du Nord (McBride-Kebert *et al.*, 2019) ont démontrée l'affinité des grands dauphins pour des zones peu profondes (<100m), ce qui est cohérent avec les résultats trouvés ici puisque ces profondeurs restent plus importantes qu'en Baie du Mont Saint Michel. De tels résultats doivent tout de même être utilisés avec précaution. En effet, ils peuvent être dépendant notre zone de prospection, qui ne dépasse que très rarement l'isobathe des 20 m. Des prospections ponctuelles au-delà de cette profondeur permettraient de valider l'absence d'individus dans ces zones.

Le seul exemple d'étude menée à une très faible profondeur est l'étude de Irvine *et al.* (1981) qui a démontré que les grands dauphins favorisent les zones avec une profondeur <4m. Or ces résultats vont à l'encontre de ceux présentés ici, les individus étudiés ne présentant pas de probabilité de présence optimale pour les profondeurs de moins de 12m.

Selon les GLM et les GAM, le facteur secondaire affectant la distribution des grands dauphins est la distance de la côte, avec les plus fortes probabilités de distribution détectées à quelques kilomètres de la côte (<3 ou 4 km). Cela peut être dû, d'une part, à l'attrait des zones peu profondes en termes de proies du fait de l'abondance et de la facilité de capture (Marini *et al.*, 2015) et, d'autre part, aux perturbations humaines plus importantes à distance de la côte (bateau de pêche, trafic maritime).

Les résultats apportés par les valeurs de marginalité et de spécialisation de l'ENFA confirment le caractère polyvalent de l'espèce et sa capacité à s'adapter à des conditions contrastées, comme l'ont déjà démontrés des travaux antérieurs (Perrin *et al.*, 2009).

Cette étude fournit de nouvelles informations sur la répartition des grands dauphins dans le golfe Normand-Breton, en montrant que les eaux situées entre l'archipel de Chausey et la Pointe du Grouin, ainsi que la zone au large de Saint Malo constitueraient un habitat optimal pour les grands dauphins, contrairement aux zones intérieures de la baie du Mont Saint Michel. En outre, le fort pouvoir prédictif du modèle pour les zones où les habitats ne sont pas adaptés confirme que l'espèce ne montrerait pas de préférence pour les zones très proches de la côte.

Bien qu'apportant des premiers éléments de réponse intéressants quant aux relations qui lient les grands dauphins à leur environnement, ces modèles n'expliquent qu'une faible partie de la distribution. L'homogénéité de la zone d'étude pourrait en être l'une des causes. En effet, la zone prospectée ne présente pas de grande variabilité en termes de gamme environnementale, en raison de sa faible échelle spatiale. On peut notamment noter qu'à l'échelle du golfe normand-breton, la concentration en chlorophylle-a est relativement homogène, ce qui explique qu'elle n'apparaît pas comme un facteur susceptible d'influencer de manière significative la distribution des grands dauphins.

Les travaux de Torres *et al.* (2008) ont démontré que les modèles basés sur des données environnementales avaient un pouvoir prédictif supérieur à celui des modèles basés sur la distribution en proies. Néanmoins, des informations complémentaires sur la distribution de certaines des proies du *Tursiops* dans le golfe normand-breton permettraient de confirmer les modèles d'habitats développés ici. De la même manière, d'autres variables comme les phases de marée ou les courants seraient intéressantes à intégrer dans des études ultérieures, afin d'approfondir la connaissance de l'utilisation de l'habitat par les grands dauphins (Methion et Díaz López, 2020; Mendes *et al.*, 2002). La zone d'étude étant sujette à une forte pression anthropique (bateau de pêche, plaisanciers, activités conchylicoles), il paraît important d'étudier également l'impact de ces activités sur le comportement des cétacés.

De même, pour étudier l'importance et la fonction écologique potentielle des habitats préférés des grands dauphins, il serait intéressant de calculer un pourcentage des plages comportementales représentatives qui chevauchent les zones de forte probabilité d'occurrence des dauphins.

En outre, dans cette étude, l'aspect temporel n'a pas été approfondi. Compléter ces résultats avec des modèles de distributions saisonniers permettrait d'apporter d'autres éléments de réponse quant aux variations de distribution des grands dauphins.

#### ✓ Intérêt de l'étude et critique des méthodes

La modélisation des relations entre les espèces et l'environnement peut fournir des informations importantes sur les processus écologiques qui déterminent la distribution des espèces, avec des implications significatives pour la conservation et la gestion. Dans ce contexte, une nouvelle approche a été conçue dans cette étude, en combinant l'utilisation de la tessellation hexagonale avec la modélisation de l'ENFA, ainsi que l'utilisation comparative des GLM, des GAM et des modèles d'ensemble.

La tessellation hexagonale a permis une adaptation adéquate à l'effort d'échantillonnage qui, avec la distance égale entre les cellules adjacentes, a permis d'élaborer des cartes éco-géographiques et d'espèce, comme cela a été vu dans d'autres études (Guisan et Zimmermann, 2000; Birch *et al.*, 2007).

En outre, l'approche de l'ENFA s'est avérée être un outil adéquat pour évaluer les habitats potentiels et la distribution d'une espèce marine dissimulée et très mobile telle que le Grand Dauphin, pour laquelle il est difficile d'obtenir des données d'absence fiables. Bien que certains auteurs aient souligné que les modèles de niche écologique ont une précision prédictive inférieure à d'autres modèles basés sur des données de présence/absence (Segurado et Araújo, 2004; Tsoar *et al.*, 2007), la validation croisée et l'indice de Boyce ont montré la robustesse du modèle ENFA et sa précision à prédire les préférences d'habitat et la distribution (MacLeod *et al.*, 2008; Praca *et al.*, 2009; Costa *et al.*, 2013). Cette étude souligne également l'importance de disposer d'un ensemble de données de présence étendu et robuste pour obtenir des résultats fiables. Cela a été confirmé par les courbes P/E utilisées pour valider le modèle, qui ont montré une plus grande précision dans la prévision des zones inadaptées à l'espèce (et dans lesquelles les dauphins n'étaient pas régulièrement observés).

La comparaison entre les GLM et les GAM a permis de montrer que les distributions prévues du grand dauphin augmentent en exactitude et en précision lorsqu'on passe du GLM au GAM. Néanmoins, l'utilisation de ces méthodes a mis en lumière la difficulté d'obtenir des données d'absence fiables pour élaborer ces modèles. La différence flagrante de prédiction entre la sélection aléatoire d'absence échantillonnée sur le terrain ( $0.675 \pm 0.02$ ) et la sélection de pseudo-absences ( $0.964 \pm 0.008$ ) montre l'importance de récolter des données d'absence selon un protocole rigoureux. La disponibilité des absences peut être critique pour évaluer correctement un habitat relativement adapté par rapport à d'autres habitats adaptés mais où les présences sont plus fréquentes car l'habitat est davantage disponible. Stockwell et Peterson (2002) ont suggéré que lorsque les données sur les absences font défaut, les modèles de distribution peuvent être améliorés en générant des pseudo-absences aléatoires à partir de zones (background) dans lesquelles les données sur les espèces sont manquantes. Cependant, cette méthode peut entraîner des données d'absence biaisées si les espèces ont une grande aire de répartition ou si les données sur la présence sont rares (Boyce *et al.*, 2002). De plus, une étude de Wisz *et al.* (2009) a démontré que la sélection de pseudo-absences influe fortement sur l'ajustement du modèle. En effet, rien n'empêche que les pseudo-absences sélectionnées au hasard tombent dans



des zones écologiquement similaires aux lieux de présence, ce qui rend l'estimation des paramètres du modèle par les algorithmes de régression très difficile (Wisz et Guisan, 2009). La mise en place d'un effet aléatoire (runs) lors de la sélection des pseudo-absences prend alors ici son sens.

En ce qui concerne la sélection du nombre de présence et du nombre d'absence dans le modèle, la méthode utilisée ici (ratio 1:1) est la plus couramment employée dans les études de distribution (Nad'ó *et al.*, 2018). Néanmoins, Nad'ó et collaborateurs (2018) ont montré que cette méthode peut être inappropriée dans le cas d'espèces difficiles à détecter et/ou rares, ce qui peut apporter des biais aux modèles générés ici. Une procédure d'échantillonnage spatialement aléatoire type GRTS ou SRE pour les absences, ainsi que la sélection d'un ratio approprié à notre population (Nad'ó *et al.*, 2018) pourraient être envisagées afin de palier à ces problématiques.

L'utilisation du package BIOMOD2 a été choisie car c'est le logiciel le plus connu permettant de créer des modèles d'ensemble, qu'il est le mieux établi au sein de la communauté SDM et qu'il est disponible gratuitement. BIOMOD2 fournit un ensemble de méthodes et d'outils pertinents pour le problème de la modélisation des distributions, comme la capacité de construire rapidement des modèles individuels et de les combiner de différentes manières. Cependant, de nombreux auteurs utilisent les paramètres par défaut de BIOMOD2 soit par manque d'expérience avec les méthodes de modélisation, soit parce qu'ils considèrent ces paramètres comme le réglage optimal. Cependant, il n'y a pas de preuve que ces réglages par défaut soient optimaux (Hao *et al.*, 2019). Même si la performance des modèles n'est pas le principal objectif de nombreuses études sur les SDM, une meilleure communication sur les détails méthodologiques serait à la fois bénéfique pour l'accumulation de connaissances dans ce domaine et pour la reproductibilité des études sur les SDM. En cela, une comparaison entre des modèles individuels générés par BIOMOD2 et des modèles générés par des packages permettant plus de manipulation des paramètres serait intéressante.

Marmion et collaborateurs (2009) ont suggéré que les ensembles peuvent améliorer la précision des prévisions de la distribution des espèces. La principale motivation derrière l'utilisation d'ensembles vient du fait que des recherches antérieures ont montré que les prévisions de la répartition des espèces peuvent varier considérablement selon les approches de modélisation. Les ensembles peuvent éviter les pièges associés à une seule approche de modélisation et l'utilisation de modèles multiples peut améliorer la capacité des chercheurs à détecter le véritable signal écologique à partir du "bruit" associé aux incertitudes des modèles individuels (Araújo et New, 2007). Dans cette étude, il est montré que les distributions des grands dauphins issues des modèles d'ensemble ne sont pas plus précises que celles générées par les modèles individuels. En effet, l'AUC des modèles d'ensemble n'était pas supérieure à celle de plusieurs algorithmes individuels, ce qui indique que dans certains cas, les modèles d'ensemble peuvent ne pas être plus utiles que les approches individuelles de modélisation. En outre, en plus de fournir des niveaux similaires de précision des distributions dans notre étude, les GLM ou les GAM présentent d'autres avantages par rapport aux ensembles. Le premier est qu'elles nécessitent moins de puissance de calcul que les modèles d'ensemble (approche plus parcimonieuse) (Crimmins *et al.*, 2013). La seconde est que ces approches se prêtent à l'interprétation du modèle par l'évaluation des courbes de réponse, ce qui facilite l'évaluation de l'occurrence le long des gradients écologiques.

Pour conclure sur le choix des modèles, il semble important de souligner que, étant basées sur des approches distinctes concernant l'ajustement aux données et la variation de la qualité des données, les méthodes de modélisation de la distribution des habitats citées ci-dessus couvriront probablement différents domaines d'application et il sera impossible d'en identifier une parmi elles comme étant universellement applicable (Elith *et al.*, 2002; Segurado et Araújo, 2004). Par conséquent, les objectifs et les hypothèses de la modélisation des habitats doivent être clairs avant d'être appliqués à des situations particulières. Les méthodes n'utilisant que des données de présence, telle que l'ENFA, conviennent à la modélisation d'habitats lorsque les données d'absences sont mauvaises ou indisponibles, ou lorsque l'objectif principal de la modélisation est d'identifier des zones globalement appropriées pour une espèce donnée. Dans le cas contraire, notamment si l'on veut prendre en compte la variabilité temporelle, les méthodologies faisant appel à des données de présence/absence doivent être privilégiées. Dans le cas de cette étude, il semble que la technique de l'ENFA soit donc la plus adaptée à nos données pour la prédiction d'habitat.

✓ Rétrospective du protocole 2019 et perspectives

En centralisant l'ensemble des données acquises par les suivis, la base créée en 2019 a permis de les exploiter sous un angle nouveau. Les premières analyses ont démontré que les données récoltées par AL LARK peuvent être utilisées pour des analyses de distribution, ce qui laisse de nombreuses perspectives d'études des données.

Une des problématiques soulevées l'an passé était un échantillonnage incomplet des données (oublis de prise de note, erreurs de tablette...). À la suite des recommandations proposées l'an dernier, une nette amélioration dans la prise de note a eu lieu (-15% de données manquantes) et l'application a été améliorée, ce qui permettra d'avoir des données de meilleure qualité pour les analyses à venir.

Concernant l'échantillonnage des données, quelques améliorations peuvent néanmoins encore être apportées. En effet, les données sont issues de prospections dont l'objectif est de trouver des groupes de grands dauphins sur l'ensemble de la zone d'étude afin de procéder à un travail de photo-identification. Or, un protocole d'échantillonnage tel que celui-ci n'est *a priori* pas le plus adapté pour des études de distribution. En effet, le caractère hétérogène de l'échantillonnage crée une importante variabilité dans l'effort de prospection de la zone et peut créer de fortes autocorrélations entre les observations lorsque les individus sont suivis. Afin d'obtenir des modèles plus précis et donc une distribution plus réaliste, il serait intéressant de réaliser des sorties dédiées, où les présences/absences seraient échantillonnées à des intervalles réguliers, en effectuant des transects permettant d'échantillonner tout le site d'étude.

A la suite de cette étude, il est apparu nécessaire de récolter des données environnementales davantage représentatives de la zone d'étude. En effet, les données issues des rasters satellitaires et les variables calculées (profondeur notamment) sont moins précises que celles que l'on peut directement échantillonner sur le terrain. À la suite de cette recommandation, un rendez-vous a été organisé avec M. Gally, fondateur de la plateforme ObsEnMer, afin d'ajouter des données sur le formulaire Al Lark de la plateforme. À l'issue de cette réunion, il a été décidé d'ajouter les variables suivantes : profondeur réelle sous le bateau, phases de marée et température de l'eau. Des informations sur le comportement des dauphins (vitesse, cap) ainsi que sur la présence de bateau (type de bateau, distance avec les dauphins, action de pêche) ont aussi été ajoutées.

De même, ce travail a permis de mettre en évidence que la recherche et l'extraction de données non récoltées par l'association demande un temps considérable. Pour remédier à ce problème, un projet a donc été déposé à l'Océan Hackathon de Saint Malo édition 2020<sup>8</sup> : créer un programme informatique (MORGAZH) permettant l'extraction automatique de données environnementales (Annexe 23).

---

<sup>8</sup> Marathon créatif de 48 heures qui se déroulera le week-end du 9 au 11 Octobre 2020 \_ <https://www.saint-malo-developpement.fr/2020/06/ocean-hackathon-saint-malo-repondez-a-lappel-a-defis-avant-le-30-juin/>  
<https://www.saint-malo-developpement.fr/2020/06/ocean-hackathon-saint-malo-repondez-a-lappel-a-defis-avant-le-30-juin/>

## Bibliographie

- Allouche, Omri, Ofer Steinitz, Dotan Rotem, Arik Rosenfeld, et Ronen Kadmon. 2008. « Incorporating distance constraints into species distribution models ». *Journal of Applied Ecology* 45 (2): 599-609. <https://doi.org/10.1111/j.1365-2664.2007.01445.x>.
- Araújo, Miguel B., et Mark New. 2007. « Ensemble forecasting of species distributions ». *Trends in Ecology & Evolution* 22 (1): 42-47. <https://doi.org/10.1016/j.tree.2006.09.010>.
- Barbet-Massin, Morgane, Frédéric Jiguet, Cécile Hélène Albert, et Wilfried Thuiller. 2012. « Selecting pseudo-absences for species distribution models: how, where and how many? » *Methods in Ecology and Evolution* 3 (2): 327-38. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>.
- Basille, Mathieu, Clément Calenge, Éric Marboutin, Reidar Andersen, et Jean-Michel Gaillard. 2008. « Assessing habitat selection using multivariate statistics: Some refinements of the ecological-niche factor analysis ». *Ecological Modelling* 211 (1): 233-40. <https://doi.org/10.1016/j.ecolmodel.2007.09.006>.
- Bandel L, 2018. Structure sociale des grands dauphins (*Tursiops truncatus*) en mer de la Manche. Mémoire de Master 2, Université Côte d'Azur et Nice Sophia Antipolis, 36 p.
- Birch, Colin P.D., Sander P. Oom, et Jonathan A. Beecham. 2007. « Rectangular and hexagonal grids used for observation, experiment and simulation in ecology ». *Ecological Modelling* 206 (3): 347-59. <https://doi.org/10.1016/j.ecolmodel.2007.03.041>.
- Bonnot-Courtois, C. (2002). Bay of Mont-Saint-Michel and the Rance Estuary: Recent Development and Evolution of the Depositional Environments (Vol. 26)
- Bonnot-courtois, Chantal. 2012. « Dynamique sédimentaire intertidale en baie du Mont-Saint-Michel entre évolution naturelle et aménagements ». XIIèmes Journées, Cherbourg. Editions Paralia, 187-222.
- Boyce, Mark S, Pierre R Vernier, Scott E Nielsen, et Fiona K.A Schmiegelow. 2002. « Evaluating resource selection functions ». *Ecological Modelling* 157 (2): 281-300. [https://doi.org/10.1016/S0304-3800\(02\)00200-4](https://doi.org/10.1016/S0304-3800(02)00200-4).
- Brodeur, Richard D, et William G Pearcy. 1992. « Effects of environmental variability on trophic interactions and food web structure in a pelagic upwelling ecosystem ». *Marine Ecology Progress Series*, 101-19.
- Cañadas, A, R Sagarminaga, et S García-Tiscar. 2002. « Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain ». *Deep Sea Research Part I: Oceanographic Research Papers* 49 (11): 2053-73. [https://doi.org/10.1016/S0967-0637\(02\)00123-1](https://doi.org/10.1016/S0967-0637(02)00123-1).
- Carpenter, G., A. N. Gillison, et J. Winter. 1993. « DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals ». *Biodiversity & Conservation* 2 (6): 667-80. <https://doi.org/10.1007/BF00051966>.
- Condet, Manon, et Violaine Dulau-Drouot. 2016. « Habitat selection of two island-associated dolphin species from the south-west Indian Ocean ». *Continental Shelf Research* 125 (août): 18-27. <https://doi.org/10.1016/j.csr.2016.06.010>.
- Costa, H, V Medeiros, E B Azevedo, et L Silva. 2013. « Evaluating ecological-niche factor analysis as a modelling tool for environmental weed management in island systems ». *Weed Research* 53 (3): 221-30. <https://doi.org/10.1111/wre.12017>.
- Cribb, Nardi, Cara E Miller, et Laurent Seuront. 2015. « Towards a standardized approach of cetacean habitat: past achievements and future directions ». *Open Journal of Marine Science* 5: 335-57.
- Crimmins, Shawn M., Solomon Z. Dobrowski, et Alison R. Mynsberge. 2013. « Evaluating ensemble forecasts of plant species distributions under climate change ». *Ecological Modelling* 266 (septembre): 126-30. <https://doi.org/10.1016/j.ecolmodel.2013.07.006>.
- David, L, et N Di-Méglio. 1999. « First results of summer movements of *Grampus griseus* (Cuvier, 1812) in the North-Western Mediterranean Sea ». *Eur. Res. Cet* 13: 189-94.
- Davies, Andrew J., Max Wisshak, James C. Orr, et J. Murray Roberts. 2008. « Predicting suitable habitat for the cold-water coral *Lophelia pertusa* (Scleractinia) ». *Deep Sea Research Part I: Oceanographic Research Papers* 55 (8): 1048-62. <https://doi.org/10.1016/j.dsr.2008.04.010>.
- Davis, Randall W, Joel G Ortega-Ortiz, Christine A Ribic, William E Evans, Douglas C Biggs, Patrick H Ressler, Robert B Cady, Robert R Leben, Keith D Mullin, et Bernd Würsig. 2002. « Cetacean habitat in the northern oceanic Gulf of Mexico ». *Deep Sea Research Part I: Oceanographic Research Papers* 49 (1): 121-42. [https://doi.org/10.1016/S0967-0637\(01\)00035-8](https://doi.org/10.1016/S0967-0637(01)00035-8).
- Dormann, Carsten F., Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, et al. 2013. « Collinearity: a review of methods to deal with it and a simulation study evaluating their performance ». *Ecography* 36 (1): 27-46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Doums–Président, M. C., Monestiez–Rapporteur, M. P., & Ridoux–Examineur, M. V. Les grands dauphins sédentaires *Tursiops truncatus* du golfe normand-breton: distribution, estimation et structure sociale de la population entre 2009 et 2013.
- Ehrhold, Axel. 1999. « Dynamique de comblement d'un bassin sédimentaire soumis à un régime mégatidal: exemple de la Baie du Mont-Saint-Michel ».
- Elith, Jane, Mark A Burgman, et Helen M Regan. 2002. « Mapping epistemic uncertainties and vague concepts in predictions of species distribution ». *Ecological Modelling* 157 (2): 313-29. [https://doi.org/10.1016/S0304-3800\(02\)00202-8](https://doi.org/10.1016/S0304-3800(02)00202-8).
- Elith, Jane, et John R Leathwick. 2009. « Species distribution models: ecological explanation and prediction across space and time ». *Annual review of ecology, evolution, and systematics* 40: 677-97.
- Farber, Oren, et Ronen Kadmon. 2003. « Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance ». *Ecological Modelling* 160 (1): 115-30. [https://doi.org/10.1016/S0304-3800\(02\)00327-7](https://doi.org/10.1016/S0304-3800(02)00327-7).
- Fernandez, M., Yesson, C., Gannier, A., Miller, P., Azevedo, J., 2018. A matter of timing: how temporal scale selection influences cetacean ecological niche modelling. *Mar. Ecol. Prog. Ser.* 595, 217–231. <https://doi.org/10.3354/meps12551>.
- Fielding, ALAN H., et JOHN F. Bell. 1997. « A review of methods for the assessment of prediction errors in conservation presence/absence models ». *Environmental Conservation* 24 (1): 38-49.

- Franklin, J. (2010). Mapping species distributions: spatial inference and prediction. Cambridge University Press.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
- Gally, F. 2016. Suivi de la population des grands dauphins sédentaires du golfe normand-breton et de la baie de Seine. Rapport de synthèse du GECC pour l'année 2014.
- Gannier, Alexandre. 1995. « Les Cétacés de Méditerranée nord-occidentale: estimation de leur abondance et mise en relation de la variation saisonnière de leur distribution avec l'écologie du milieu ».
- Gaudin, F. (2017) . Effets du changement climatique sur la distribution de la macrofaune benthique en Manche. Océanographie. Université Pierre et Marie Curie - Paris VI, 2017. Français. (NNT : 2017PA066182). (tel-01661329)
- Gohin, F., Saulquin, B., & Bryere, P. (2010). Atlas de la Température, de la concentration en Chlorophylle et de la Turbidité de surface du plateau continental français et de ses abords de l'Ouest européen.
- Gnone, Guido, Michela Bellingeri, Frank Dhermain, Franck Dupraz, Silvio Nuti, Davide Bedocchi, Aurelie Moulins, et al. 2011. « Distribution, abundance, and movements of the bottlenose dolphin (*Tursiops truncatus*) in the Pelagos Sanctuary MPA (north-west Mediterranean Sea) ». *Aquatic Conservation: Marine and Freshwater Ecosystems* 21 (4): 372-88. <https://doi.org/10.1002/aqc.1191>.
- Grenouillet, Gael, Laetitia Buisson, Nicolas Casajus, et Sovan Lek. 2011. « Ensemble modelling of species distribution: the effects of geographical and environmental ranges ». *Ecography* 34 (1): 9-17. <https://doi.org/10.1111/j.1600-0587.2010.06152.x>.
- Guisan, Antoine, Thomas C Edwards Jr, et Trevor Hastie. 2002. « Generalized linear and generalized additive models in studies of species distributions: setting the scene ». *Ecological modelling* 157 (2-3): 89-100.
- Guisan, Antoine, et Wilfried Thuiller. 2005. « Predicting species distribution: offering more than simple habitat models ». *Ecology letters* 8 (9): 993-1009.
- Guisan, Antoine, et Niklaus E Zimmermann. 2000. « Predictive habitat distribution models in ecology ». *Ecological modelling* 135 (2-3): 147-86.
- Hamazaki, Toshihide. 2002. « Spatiotemporal prediction models of cetacean habitats in the mid-western North Atlantic Ocean (from Cape Hatteras, North Carolina, USA to Nova Scotia, Canada) ». *Marine Mammal Science* 18 (4): 920-39.
- Hao, T, Elith, J, Guillera-Aroita, G, Lahoz-Monfort, J.J. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. *Divers Distrib.* 2019; 25: 839–852. <https://doi.org/10.1111/ddi.12892>
- Hastie, T. J., & Tibshirani, R. J. (1990). Generalized additive models (Vol. 43). CRC press.
- Hastie, G. D., B. Wilson, L. J. Wilson, K. M. Parsons, et P. M. Thompson. 2004. « Functional mechanisms underlying cetacean distribution patterns: hotspots for bottlenose dolphins are linked to foraging ». *Marine Biology* 144 (2): 397-403. <https://doi.org/10.1007/s00227-003-1195-4>.
- Hastie, Gordon D., René J. Swift, George Slesser, Paul M. Thompson, et William R. Turrell. 2005. « Environmental models for predicting oceanic dolphin habitat in the Northeast Atlantic ». *ICES Journal of Marine Science* 62 (4): 760-70. <https://doi.org/10.1016/j.icesjms.2005.02.004>.
- Hirzel. 2008. « Habitat suitability modelling and niche theory ». *Journal of Applied Ecology* 45 (5): 1372-81. <https://doi.org/10.1111/j.1365-2664.2008.01524.x>.
- Hirzel, A.H., V. Helfer, et F. Metral. 2001. « Assessing habitat-suitability models with a virtual species ». *Ecological Modelling* 145 (2): 111-21. [https://doi.org/10.1016/S0304-3800\(01\)00396-9](https://doi.org/10.1016/S0304-3800(01)00396-9).
- Hirzel, Alexandre, et Antoine Guisan. 2002. « Which is the optimal sampling strategy for habitat suitability modelling ». *Ecological Modelling* 157 (2): 331-41. [https://doi.org/10.1016/S0304-3800\(02\)00203-X](https://doi.org/10.1016/S0304-3800(02)00203-X).
- Hirzel, Alexandre H, Jacques Hausser, Daniel Chessel, et Nicolas Perrin. 2002. « Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? » *Ecology* 83 (7): 2027-36.
- Hirzel, A. H., & Arlettaz, R. (2003). Modeling habitat suitability for complex species distributions by environmental-distance geometric mean. *Environmental management*, 32(5), 614-623.
- Hirzel, Alexandre H., Gwenaëlle Le Lay, Véronique Helfer, Christophe Randin, et Antoine Guisan. 2006. « Evaluating the ability of habitat suitability models to predict species presences ». *Predicting Species Distributions* 199 (2): 142-52. <https://doi.org/10.1016/j.ecolmodel.2006.05.017>.
- Hirzel, BERTRAND POSSE, PIERRE-ALAIN OGGIER, YVON CRETENAND, CHRISTIAN GLENZ, et RAPHAËL ARLETTAZ. 2004. « Ecological requirements of reintroduced species and the implications for release policy: the case of the bearded vulture ». *Journal of Applied Ecology* 41 (6): 1103-16. <https://doi.org/10.1111/j.0021-8901.2004.00980.x>.
- Hutchinson, G.E., 1957. Concluding remarks. *Cold Spring Harbor Symp. Quant. Biol.* 22, 415–427.
- Irvine, A. B., Scott, M. D., Wells, R. S., & Kaufmann, J. H. (1981). Movements and activities of the Atlantic bottlenose dolphin, *Tursiops truncatus*, near Sarasota, Florida. *Fishery bulletin*, 79(4), 671-688.
- Kermorvant Claire(1), D'Amico Frank(1), Bru Noëlle(1) , Caill-Milly Nathalie(2), Poster, De nouveaux packages pour sélectionner des points d'échantillonnage spatialement équilibrés sous R
- Laffaille, P., E. Feunteun, et J.-C. Lefeuvre. 2000. « Composition of Fish Communities in a European Macrotidal Salt Marsh (the Mont Saint-Michel Bay, France) ». *Estuarine, Coastal and Shelf Science* 51 (4): 429-38. <https://doi.org/10.1006/ecss.2000.0675>.
- Le Mao, Patrick, Pierre-Yves Pasco, et Sébastien Provost. 2006. « Consommation de la macro-faune invertébrée benthique par les oiseaux d'eau en baie du Mont-Saint-Michel ». *Alauda* 74 (1): 23-36.
- Lima, Steven L, et Lawrence M Dill. 1990. « Behavioral decisions made under the risk of predation: a review and prospectus ». *Canadian journal of zoology* 68 (4): 619-40.
- Li, J., & Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.
- López, Bruno Díaz, et Séverine Methion. 2017. « The impact of shellfish farming on common bottlenose dolphins' use of habitat ». *Marine Biology* 164 (4): 83.
- López, Bruno Díaz, et Séverine Methion. 2018. « Does interspecific competition drive patterns of habitat use and

- relative density in harbour porpoises? » *Marine Biology* 165 (5): 92.
- Louis, Marie, François Gally, Christophe Barbraud, Julie Béseau, Paul Tixier, Benoit Simon-Bouhet, Kevin Le Rest, et Christophe Guinet. 2015. « Social structure and abundance of coastal bottlenose dolphins, *Tursiops truncatus*, in the Normano-Breton Gulf, English Channel ». *Journal of mammalogy* 96 (3): 481-93.
- Louis, Marie, Amélia Viricel, Tamara Lucas, Hélène Peltier, Eric Alfonsi, Simon Berrow, Andrew Brownlow, et al. 2014. « Habitat-Driven Population Structure of Bottlenose Dolphins, *Tursiops Truncatus*, in the North-East Atlantic ». *Molecular Ecology* 23 (4): 857-74. <https://doi.org/10.1111/mec.12653>.
- Lu, G.Y., Wong, D.W., 2008. An adaptive inverse-distance weighting spatial interpolation technique. *Comput. Geosci.* 34, 1044–1055. <https://doi.org/10.1016/j.cageo.2007.07.010>.
- MacLeod, Colin D., Laura Mandleberg, Caroline Schweder, Sarah M. Bannon, et Graham J. Pierce. 2008. « A comparison of approaches for modelling the occurrence of marine animals ». In *Essential Fish Habitat Mapping in the Mediterranean*, édité par Vasilis D. Valavanis, 21-32. Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-1-4020-9141-4\\_3](https://doi.org/10.1007/978-1-4020-9141-4_3).
- Manel, Stéphanie, Jean-Marie Dias, et Steve J. Ormerod. 1999. « Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird ». *Ecological Modelling* 120 (2): 337-47. [https://doi.org/10.1016/S0304-3800\(99\)00113-1](https://doi.org/10.1016/S0304-3800(99)00113-1).
- Marini, C., F. Fossa, C. Paoli, M. Bellingeri, G. Gnone, et P. Vassallo. 2015. « Predicting bottlenose dolphin distribution along Liguria coast (northwestern Mediterranean Sea) through different modeling techniques and indirect predictors ». *Journal of Environmental Management* 150 (mars): 9-20. <https://doi.org/10.1016/j.jenvman.2014.11.008>.
- Marmion, Mathieu, Miia Parviainen, Miska Luoto, Risto K. Heikkinen, et Wilfried Thuiller. 2009. « Evaluation of consensus methods in predictive species distribution modelling ». *Diversity and Distributions* 15 (1): 59-69. <https://doi.org/10.1111/j.1472-4642.2008.00491.x>.
- McBride-Kebert, S., Taylor, J. S, Wilkinson, K. A, Lyn, H., Moore, F. R, Sacco, D. F, et al. (2019). Common bottlenose dolphin, *Tursiops truncatus*, seasonal habitat use and associations with habitat characteristics in Roanoke Sound, North Carolina. *International Journal of Comparative Psychology*, 32. Retrieved from <https://escholarship.org/uc/item/2fn0n4wf>
- McCullagh, P., & Nelder, J. A. (1989). *Monographs on statistics and applied probability. Generalized linear models*, 37.
- Ménesguen A, et Hoch T. 1997. « Modelling the biogeochemical cycles of elements limiting primary production in the English Channel: I. Role of thermohaline stratification ». *Marine Ecology Progress Series* 146: 173-88.
- Ménesguen, Alain, Philippe Cugier, Sophie Loyer, Alice Vanhoutte-Brunier, Thierry Hoch, Jean-François Guillaud, et Francis Gohin. 2007. « Two-or three-layered box-models versus fine 3D models for coastal ecological modelling? A comparative study in the English Channel (Western Europe) ». *Journal of Marine Systems* 64 (1-4): 47-65.
- Methion, Séverine, et Bruno Díaz López. 2020. « Individual foraging variation drives social organization in bottlenose dolphins ». *Behavioral Ecology* 31 (1): 97-106. <https://doi.org/10.1093/beheco/arz160>.
- Migné, A., N. Spilmont, G. Boucher, L. Denis, C. Hubas, M.-A. Janquin, M. Rauch, et D. Davoult. 2009. « Annual budget of benthic production in Mont Saint-Michel Bay considering cloudiness, microphytobenthos migration, and variability of respiration rates with tidal conditions ». *Continental Shelf Research* 29 (19): 2280-85. <https://doi.org/10.1016/j.csr.2009.09.004>.
- Moura, André E, Neftalí Sillero, et Ana Rodrigues. 2012. « Common dolphin (*Delphinus delphis*) habitat preferences using data from two platforms of opportunity ». *Acta oecologica* 38: 24-32.
- Nad'ó L, Kaňuch P (2018) Why sampling ratio matters: Logistic regression and studies of habitat use. *PLoS ONE* 13(7): e0200742. <https://doi.org/10.1371/journal.pone.0200742>
- Natoli, Ada, Alexei Birkun, Alex Aguilar, Alfredo Lopez, et A. Rus Hoelzel. 2005. « Habitat structure and the dispersal of male and female bottlenose dolphins (*Tursiops truncatus*) ». *Proceedings of the Royal Society B: Biological Sciences* 272 (1569): 1217-26. <https://doi.org/10.1098/rspb.2005.3076>.
- Paradell, Oriol Giralt, Bruno Díaz López, et Séverine Methion. 2019. « Modelling common dolphin (*Delphinus delphis*) coastal distribution and habitat use: Insights for conservation ». *Ocean & Coastal Management* 179: 104836.
- PELAGOS. 2007. Code de bonne conduite pour l'observation des cétacés
- Pérez-Jorge, Sergi, Thalia Pereira, Chloe Corne, Zeno Wjitten, Mohamed Omar, Jillo Katello, Mark Kinyua, Daniel Oro, et Maite Louzao. 2015. « Can Static Habitat Protection Encompass Critical Areas for Highly Mobile Marine Top Predators? Insights from Coastal East Africa ». *PLOS ONE* 10 (7): e0133265. <https://doi.org/10.1371/journal.pone.0133265>.
- Perri et Petiau, 2019. Rapport d'activités scientifiques de l'association AL LARK
- Perrin, William F, Bernd Würsig, et JGM Thewissen. 2009. *Encyclopedia of marine mammals*. Academic Press.
- Phillips, Steven J., Robert P. Anderson, et Robert E. Schapire. 2006. « Maximum entropy modeling of species geographic distributions ». *Ecological Modelling* 190 (3): 231-59. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Pikesley, Stephen K., Annette C. Broderick, Daniel Cejudo, Michael S. Coyne, Matthew H. Godfrey, Brendan J. Godley, Pedro Lopez, et al. 2015. « Modelling the niche for a marine vertebrate: a case study incorporating behavioural plasticity, proximate threats and climate change ». *Ecography* 38 (8): 803-12. <https://doi.org/10.1111/ecog.01245>.
- Pirotta, Enrico, Jason Matthiopoulos, Monique MacKenzie, Lindsay Scott-Hayward, et Luke Rendell. 2011. « Modelling sperm whale habitat preference: a novel approach combining transect and follow data ». *Marine Ecology Progress Series* 436: 257-72.
- Planque, Benjamin, Christophe Loots, Pierre Petitgas, ULF Lindstrøm, et Sandrine Vaz. 2011. « Understanding what controls the spatial distribution of fish populations using a multi-model approach ». *Fisheries Oceanography* 20 (1): 1-17.
- Praca, Emilie, Alexandre Gannier, Krishna Das, et Sophie Laran. 2009. « Modelling the habitat suitability of cetaceans: Example of the sperm whale in the northwestern Mediterranean Sea ». *Deep Sea Research Part I: Oceanographic Research Papers* 56 (4): 648-57. <https://doi.org/10.1016/j.dsr.2008.11.001>.
- Praca, E., & Gannier, A. (2008). Ecological niches of three teuthophageous odontocetes in the northwestern Mediterranean Sea.

- Redfern, JV, Megan C Ferguson, EA Becker, KD Hyrenbach, Caroline Good, Jay Barlow, K Kaschner, Mark F Baumgartner, Karin A Forney, et Lisa T Ballance. 2006. « Techniques for cetacean-habitat modeling ». *Marine Ecology Progress Series* 310: 271-95.
- Reutter, Brigitte A., V. Helfer, A. H. Hirzel, et P. Vogel. 2003. « Modelling habitat-suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps ». *Journal of Biogeography* 30 (4): 581-90. <https://doi.org/10.1046/j.1365-2699.2003.00855.x>.
- Riul P, Targino CH, Júnior LAC, Creed JC, Horta PA, et Costa GC. 2013. « Invasive potential of the coral *Tubastraea coccinea* in the southwest Atlantic ». *Marine Ecology Progress Series* 480: 73-81.
- Rodriguez-Ferrer, Grisel, Juan J. Cruz-Motta, Nikolaos V. Schizas, et Richard S. Appeldoorn. 2020. « Modelling distribution of the common bottlenose dolphin, *Tursiops truncatus* off the southwest coast of Puerto Rico ». *Journal of Marine Systems* 210 (octobre): 103371. <https://doi.org/10.1016/j.jmarsys.2020.103371>.
- Ross, GJB. 1977. « The taxonomy of bottlenose dolphins *Tursiops* species in South African waters, with notes on their biology ». *Ann. Cape Prov. Mus. Nat. Hist.* 11: 135-94.
- Santos, MB, R Fernández, A López, JA Martínez, et GJ Pierce. 2007. « Variability in the diet of bottlenose dolphin, *Tursiops truncatus*, in Galician waters, north-western Spain, 1990–2005 ». *Journal of the Marine Biological Association of the United Kingdom* 87 (1): 231-41.
- Scales, Kylie L., Peter I. Miller, Simon N. Ingram, Elliott L. Hazen, Steven J. Bograd, et Richard A. Phillips. 2016. « Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models ». *Diversity and Distributions* 22 (2): 212-24. <https://doi.org/10.1111/ddi.12389>.
- Segurado, Pedro, et Miguel B. Araújo. 2004. « An evaluation of methods for modelling species distributions ». *Journal of Biogeography* 31 (10): 1555-68. <https://doi.org/10.1111/j.1365-2699.2004.01076.x>.
- Skov, Henrik, Elizabeth Humphreys, Stefan Garthe, Kerstin Geitner, David Grémillet, Keith C. Hamer, Janos Hennenke, Hjalte Parner, et Sarah Wanless. 2008. « Application of habitat suitability modelling to tracking data of marine animals as a means of analyzing their feeding habitats ». *Ecological Modelling* 212 (3): 504-12. <https://doi.org/10.1016/j.ecolmodel.2007.11.006>.
- Smith, Jennifer E, Joseph M Kolowski, Katharine E Graham, Stephanie E Dawes, et Kay E Holekamp. 2008. « Social and ecological determinants of fission–fusion dynamics in the spotted hyaena ». *Animal Behaviour* 76 (3): 619-36.
- Sónia Mendes, William Turrell, Thomas Lütkebohle, et Paul Thompson. 2002. « Influence of the tidal cycle and a tidal intrusion front on the spatio-temporal distribution of coastal bottlenose dolphins ». *Marine Ecology Progress Series* 239: 221-29.
- Spyrakos, Evangelos, Tania C. Santos-Diniz, Gema Martinez-Iglesias, Jesus M. Torres-Palenzuela, et Graham J. Pierce. 2011. « Spatiotemporal patterns of marine mammal distribution in coastal waters of Galicia, NW Spain ». *Hydrobiologia* 670 (1): 87. <https://doi.org/10.1007/s10750-011-0722-4>.
- Stockwell, David R.B, et A.Townsend Peterson. 2002. « Effects of sample size on accuracy of species distribution models ». *Ecological Modelling* 148 (1): 1-13. [https://doi.org/10.1016/S0304-3800\(01\)00388-X](https://doi.org/10.1016/S0304-3800(01)00388-X).
- Swets, JA. 1988. « Measuring the accuracy of diagnostic systems ». *Science* 240 (4857): 1285. <https://doi.org/10.1126/science.3287615>.
- Tabor, Gary M, et A Alonso Aguirre. 2004. « Ecosystem health and sentinel species: adding an ecological element to the proverbial “canary in the mineshaft” ». *EcoHealth* 1 (3): 226-28.
- Thuiller, Wilfried, Bruno Lafourcade, Robin Engler, et Miguel B. Araújo. 2009. « BIOMOD – a platform for ensemble forecasting of species distributions ». *Ecography* 32 (3): 369-73. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>.
- Torres, Leigh G, Andrew J Read, et Patrick Halpin. 2008. « Fine-scale habitat modeling of a top marine predator: Do prey data improve predictive capacity ». *Ecological Applications* 18 (7): 1702-17.
- Trigui, Rima Jihane. 2009. « Influence des facteurs environnementaux et anthropiques sur la structure et le fonctionnement des peuplements benthiques du Golfe Normano-Breton ».
- Tsoar, Asaf, Omri Allouche, Ofer Steinitz, Dotan Rotem, et Ronen Kadmon. 2007. « A comparative evaluation of presence-only methods for modelling species distribution ». *Diversity and Distributions* 13 (4): 397-405. <https://doi.org/10.1111/j.1472-4642.2007.00346.x>.
- Weir, Caroline R, Colin D MacLeod, et Graham J Pierce. 2012. « Habitat preferences and evidence for niche partitioning amongst cetaceans in the waters between Gabon and Angola, eastern tropical Atlantic ». *Marine Biological Association of the United Kingdom. Journal of the Marine Biological Association of the United Kingdom* 92 (8): 1735.
- Wells, Randall S, Howard L Rhinehart, Larry J Hansen, Jay C Sweeney, Forrest I Townsend, Rae Stone, David R Casper, Michael D Scott, Aleta A Hohn, et Teri K Rowles. 2004. « Bottlenose dolphins as marine ecosystem sentinels: developing a health monitoring system ». *EcoHealth* 1 (3): 246-54.
- Wells, Randall S, et Michael D Scott. 1999. « Bottlenose dolphin *tursiops truncatus* (montagu, 1821) ». *Handbook of marine mammals: the second book of dolphins and porpoises* 6: 137-82.
- Wilson, Matthew, Konstantinos M. Andreadis, Douglas E. Alsdorf, Dennis P. Lettenmaier, Delwyn Moller, et Michael Durand. 2008. « Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model ». *Geophysical Research Letters* 35 (20). <https://doi.org/10.1029/2008GL034150>.
- Wisz, Mary S., et Antoine Guisan. 2009. « Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data ». *BMC Ecology* 9 (1): 8. <https://doi.org/10.1186/1472-6785-9-8>.
- Würsig, BERND. 1978. « OCCURRENCE AND GROUP ORGANIZATION OF ATLANTIC BOTTLENOSE PORPOISES (*TURSIOPS TRUNCATUS*) IN AN ARGENTINE BAY ». *The Biological Bulletin* 154 (2): 348-59. <https://doi.org/10.2307/1541132>.
- Zaniewski, A.Elizabeth, Anthony Lehmann, et Jacob McC Overton. 2002. « Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns ». *Ecological Modelling* 157 (2): 261-80. [https://doi.org/10.1016/S0304-3800\(02\)00199-0](https://doi.org/10.1016/S0304-3800(02)00199-0).

# Annexes

## Annexe 1. Calendrier du stage

Liste	2020										
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	
Bibliographie			Bibliographie								
Mise en Forme des Données 2016 -2019			Mise en Forme des Donne								
Reformatage des données				Reformatage des données							
Cartographie descriptive (EP,AR,TR,Site d'étude)				Cartog							
Statistiques descriptives 2016 - 2019				Statistiq							
Récupération et calcul de nouvelles variables					Récupération et calcul de nouve						
Modélisation ENFA/GLM/GAM							Modélisation ENFA/GLM				
Rédaction du rapport							Rédaction du rapport				

## Annexe 2. Répartition du travail

Intervenant	Idee originale	Collecte des données	Bibliographie	Mise en forme des données 2016	Mise en forme des données 2019	Analyses descriptives	Cartographie	Reformatage des données	Récupération de nouvelles variables - Calcul de nouvelles variables	Création des cartes d'espèces et des cartes VEG	Analyse statistique Modélisation	Cartographie d'habitat	Rédaction
Principal	OO		OO	OO	0	OO	OO	OO	OO	OO	OO	OO	OO
Secondaires	EP-MP	HQG-GG-EP-MP-OO-A		LH					EP (marégraphe + intervalles d'absences de 10min)	JB (aide à la reflexion)			Correction : EP

OO: Oihana Olhasque

LH : Lucas His (Stagiaire BTS)

MP : Morgane Perri (Biologiste Marine)

EP : Estelle Petiau (Biologiste Marine, Maître de Stage)

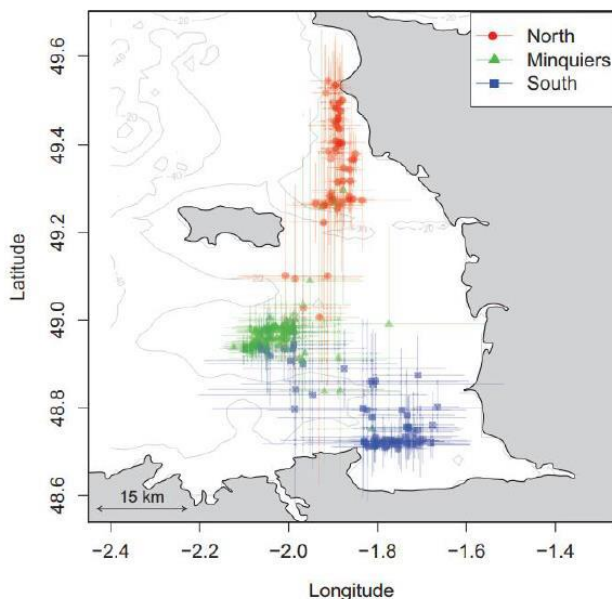
A : Adhérents

GG : Gaël Gautier (Directeur)

HQG: Hélène Quévieux Grimont (CDI)

JB: Jeremy Dolbois (Archéologue, Adhérent de l'association)

Annexe 3. Carte des observations des *Tursiops truncatus* dans le golfe normand-breton. Les codes de couleurs et de symboles indiquent le groupe social (sous-population) de chaque individu (source : Louis et al., 2015).



Annexe 4. Tableau récapitulatif des données environnementales disponibles et de leur format, pour les années de 2016 à 2019

		<b>TapForm (2016-17)</b>	<b>OBSenMER (2018-19)</b>
<b>Données météorologiques</b>	<b>Couverture nuageuse</b>	Ecriture libre (texte)	Pourcentage
	<b>Brume</b>	Ecriture libre (texte)	Oui/non
	<b>Précipitation</b>	Ecriture libre (texte)	Oui/non
	<b>Visibilité</b>	En miles + adjectif	Liste de choix (adjectif)
	<b>Force du vent</b>	Beaufort	Nœuds
	<b>Origine du vent</b>	Cardinal	Cardinal
<b>Données sur la mer</b>	<b>Hauteur de houle</b>	Non disponible	Centimètres
	<b>Direction de la houle</b>	Non disponible	Cardinal
	<b>Hauteur de vagues</b>	Non disponible	Centimètres
	<b>Etat de la mer</b>	Ecriture libre (texte)	Liste de choix selon Beaufort (adjectif)
<b>Données sur la marée</b>	<b>Coefficient de marée</b>	De 20 à 120	Non disponible
	<b>Heures d'étale</b>	Horaire (00 :00 :00)	Non disponible
	<b>Mouvement</b>	Liste de choix (MM-MD)	Non disponible



Annexe 5. Fiches terrain

a) Fiches Terrain TapForm (2016-2017)

Pad 14:34 97%

Enregistrements Morgane - Suivi mammifère...

▼ Infos sortie:

Date	13 juillet 2017
Sortie n°:	51
Heure du début de sortie:	10:00
Heure de fin de sortie:	12:51
Responsable fiche de données:	Morgane

▼ Météo / Marée:

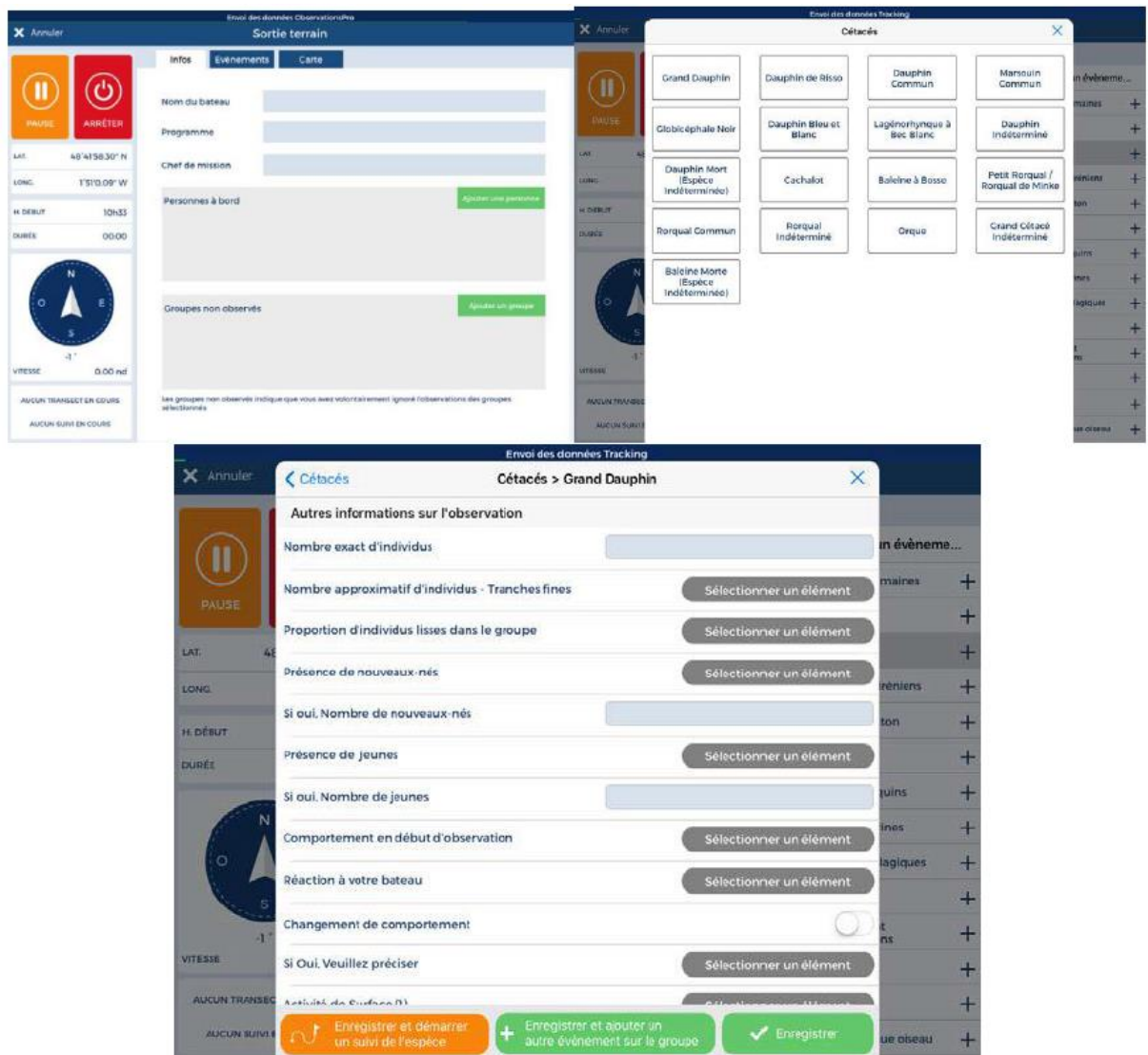
Heure de basse mer:	17:11
Heure de pleine mer:	10:31
Coefficient marée:	74
Mouvement de la marée:	Marée descendante
Direction principale du vent:	Nord-ouest
Force du vent (Beaufort):	2
Force du vent (noeuds):	5
Météorologie:	Ensoleillée
Visibilité (en miles + adjectif):	>10
Remarques (météo, marée, mer, embarcation, ...):	

▼ Informations observation(s)

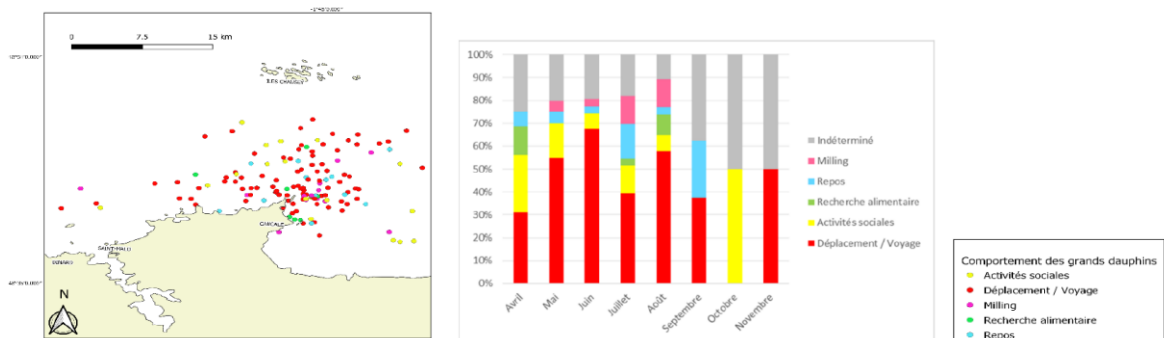
Observation n°1:	<input checked="" type="checkbox"/>
Heure du début de l'observation n°1:	11:06
Position du début de l'observation n°1:	48.752938,-1.865649
1 - Indice de présence:	Appel VHF
1 - Si appel, précisez:	Par inga
1 - Espèce:	Grand dauphin
1 - Nombre d'individus:	40
1 - Structure du groupe:	Sous-groupes
1 - Activité:	Déplacement lent
1 - Si autre, précisez:	
1 - Petit(s):	<input type="radio"/>
1 - Changement(s) (activités, structure, cap, vitesse, position, heure, ...)	
Heure de fin d'observation n°1:	12:07
Position de fin d'observation n°1:	48.752240,-1.865469
1 - Photo(s) / Vidéo(s):	<input checked="" type="checkbox"/>
1 - Autre(s) bateau(x)	<input type="radio"/>

85 de 106

b) Fiches terrain ObsenEnMER



Annexe 6. a) Carte de la distribution des comportements dans la zone b) Type de comportement en fonction du mois

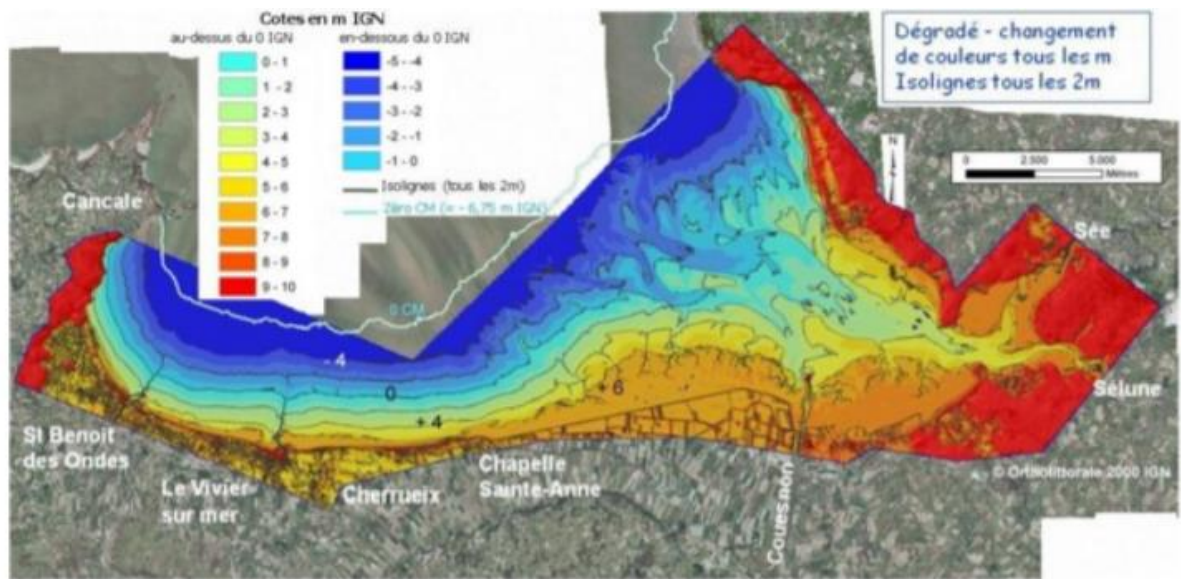


Annexe 7. Dictionnaire de la base de données créée en 2019 pendant le stage de M1

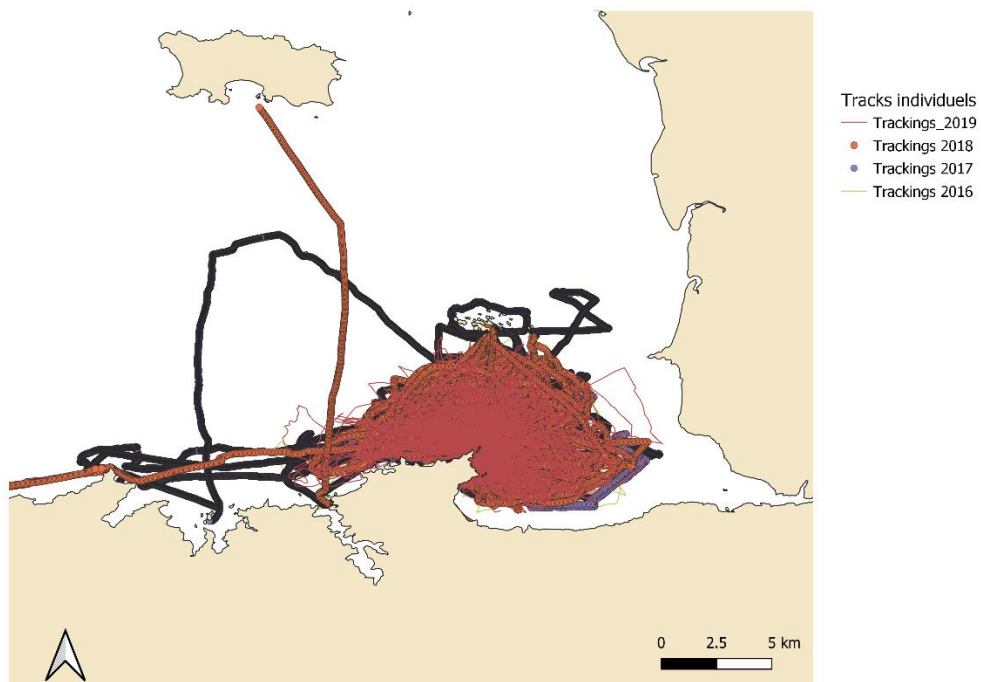
Feuille Sortie	Feuille Observations	Feuille Data Cétacés Gdauphins
Numéro de Sortie	Sortie	Sortie
Date de la sortie (jj/mm/aaaa)	Date de la sortie	Date de la sortie
Observations (Oui/Non)	Numéro d'échantillonnage	Numéro d'échantillonnage
Heure début sortie (00:00:00)	Heure de début	Observateur (chef de mission)
Heure fin sortie (00:00:00)	Heure de fin	Heure de début
Temps de pause (00:00:00)	Durée d'observation	Heure de fin
Moment de la journée	Latitude début obs	Durée d'observation
Effort d'observation (00:00:00, Formule)	Longitude début obs	Médiane Nb individus
Latitude début d'effort (degré minute secondes)	Latitude fin obs	Nombre exact d'individus
Longitude début d'effort (degré minute secondes)	Longitude fin obs	Nombre approximatif d'individus - Tranches fines
Latitude fin d'effort (degré minute secondes)	Observateur (chef de mission)	Proportion d'individus lisses dans le groupe
Longitude fin d'effort (degré minute secondes)	Nom du bateau	Présence de nouveaux-nés
Observateur (chef de mission)	Coefficient de marée	Nombre de nouveaux-nés
Nom du bateau	Etat de la Marée	Présence de jeunes
Distance parcourue (km)	Heure PM	Nombre de jeunes
Coefficient de marée	Heure BM	Comportement en début d'observation
Etat marée début sortie (MD-MM-PM-BM)		Réaction au bateau
Etat marée fin sortie		Changement de comportement
Heure météo		Description
Latitude météo		Activité de Surface
Longitude météo		Vitesse de nage
Force du vent (nœuds)		Structure du groupe
Origine du vent (cardinal)		Nombre de bateaux autour du groupe (sauf vous)
Visibilité (liste de choix)		Prise de photographies
Couverture nuageuse (%)		Représentation photo du groupe
Brume (Oui/Non)		Prise de vidéos
Précipitation (Oui/Non)		Code PHOTO ID (nom du dossier)
Etat de la mer (échelle de beaufort)		Etat de l'animal / des animaux
Degré de Beaufort		Animal Vivant ou Mort?
Direction houle		Code de décomposition
Hauteur houle (cm)		Présence d'activités humaines + Code
Hauteur vagues (cm)		Autre
Tracking (suite de lat-long-heure)		Présence de déchets
		Type de déchet
		Individus identifiés

Feuilles Cétacés Marsouins - Dauphins Commun- Risso Idem Gdauphins	Feuille Oiseaux	Feuille Poissons	Feuille Déchets	Feuille Activités humaines	Feuille papillons	Feuille macroplankton
Sortie	Sortie	Sortie	Sortie	Sortie	Sortie	Sortie
Date de la sortie	Date de la sortie	Date de la sortie	Date de la sortie	Date de la sortie	Date de la sortie	Date de la sortie
Numéro d'échantillonnage	Numéro d'échantillonnage	Numéro d'échantillonnage	Numéro d'échantillonnage	Numéro d'échantillonnage	Numéro d'échantillonnage	Numéro d'échantillonnage
Heure de début	Heure de début	Heure de début	Heure de début	Heure de début	Heure de début	Heure de début
Heure de fin	Heure de fin	Heure de fin	Heure de fin	Heure de fin	Heure de fin	Heure de fin
Durée d'observation	Durée d'observation	Durée d'observation	Durée d'observation	Durée d'observation	Durée d'observation	Durée d'observation
Latitude	Latitude	Latitude	Latitude	Latitude	Latitude	Latitude
Longitude	Longitude	Longitude	Longitude	Longitude	Longitude	Longitude
Espèce	Identification de l'espèce	TYPE	Description	Nombre d'individus	Identification de l'espèce	
Nombre exact d'individus	Nombre exact d'individus	Taille macrodéchets (cm)	Nombre exact observé		Nombre exact d'individus	
Nombre approximatif d'individus - tranches larges	Nombre approximatif d'individus - tranches larges	Nombre	Activité observée		Nombre d'individus dans le champ de vision	
Comportement	Comportement en début d'observation	Amas	Distance & Position		Classe de couleurs	
Axe de vol / cap	Etat de l'animal / des animaux	Commentaires	Distance exacte en mètres		Autres commentaires	
Cap Variable	Animal Vivant ou Mort		Distance approximative			
Age			Réticule (= Graduations des jumelles paramétriques)			
Réaction au bateau			gisement (direction de la cible par rapport au cap du bateau)			
Plumage			relevement (direction de la cible par rapport au nord, voir jumelles paramétriques)			
Mue						
Distance sampling						
Distance exacte (mètres)						
Distance +/- 300m						
Hauteur de vol						
Réticule / graduation						
gisement						
Etat de l'animal / des animaux						

Annexe 8. MNT de la zone intertidale extrait du LIDAR 2002.



Annexe 9. Trackings GPS, sous forme de polygones (2019) et de suite de points (2017)



#### A- Algorithme Médian.

Pour le calculer, l'amplitude occupée par l'espèce dans chaque facteur est divisée en 25 classes, de telle façon à ce que la médiane va séparer exactement 2 classes. Pour chaque point de l'espace environnemental, le nombre d'observations dans la même classe est compté et également les observations qui sont dans n'importe quelle classe de l'autre côté de la médiane. Ce nombre est multiplié par deux puis divisé par le nombre total d'observations afin d'opérer à une normalisation. Un point appartenant à l'une des deux classes directement adjacentes à la médiane aura donc une valeur de 1 et un point situé à l'extérieur de la distribution observée aura une valeur de 0. Enfin, l'indice d'habitat probable pour ce point est calculé en faisant une moyenne pondérée de ces valeurs pour chaque facteur retenu (chaque dimension), la pondération provenant de la quantité d'information expliquée pour chacun d'entre eux. L'algorithme part de l'hypothèse que le meilleur habitat se situe au niveau de la médiane de la distribution de l'espèce sur chaque facteur (Fig. ).

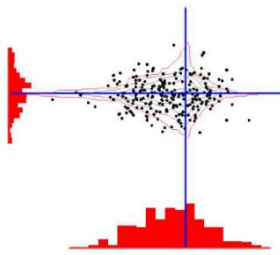


Fig. ci contre : Représentation de l'enveloppe médiane (Hirzel et al. 2012, Hirzel 2005). L'enveloppe est définie par la fréquence de distribution et la médiane.

Cet algorithme bénéficie également de deux options. D'une part l'utilisateur peut demander d'ajuster à la fréquence globale (Fig.), ce qui change la façon dont l'histogramme est construit. Le nombre d'observations (points d'occurrence) est compté, mais il est ensuite divisé par le nombre de cellules de la zone d'étude qui sont dans la même classe. Cela signifie que les habitats les plus rares (les moins présents dans la zone d'étude) ont une plus forte valeur. La médiane est ensuite recalculée. La deuxième option consiste à utiliser les valeurs extrêmes (« extreme optimum ») (Fig.) pour calculer l'optimum pour le facteur de marginalité : plutôt que de se baser sur le fait que l'optimum repose sur la médiane, l'algorithme part du principe que l'optimum repose sur le minimum ou le maximum du facteur de marginalité, selon la forme de l'histogramme. Cela résout le problème rencontré avec les espèces vivant dans des habitats rares ou dont l'habitat optimal n'existe plus. Il sera utile pour l'extrapolation lors de la mise en place de scénarios de gestion. **La première option sera donc utilisée si l'espèce vit dans des habitats rares, la deuxième si l'espèce vit dans des habitats en cours de disparition.**

#### B. Algorithme de la distance géométrique moyenne

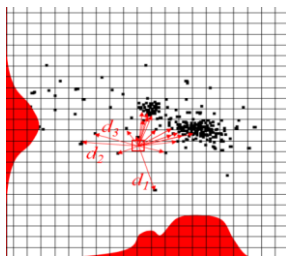
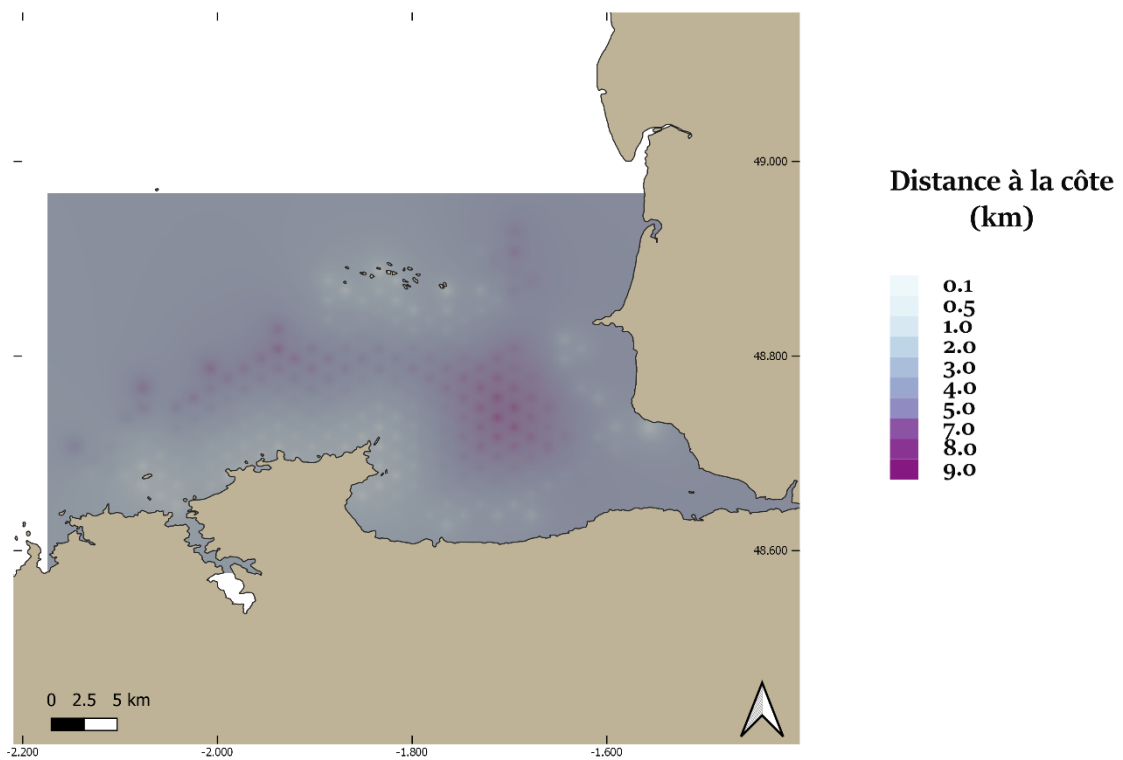
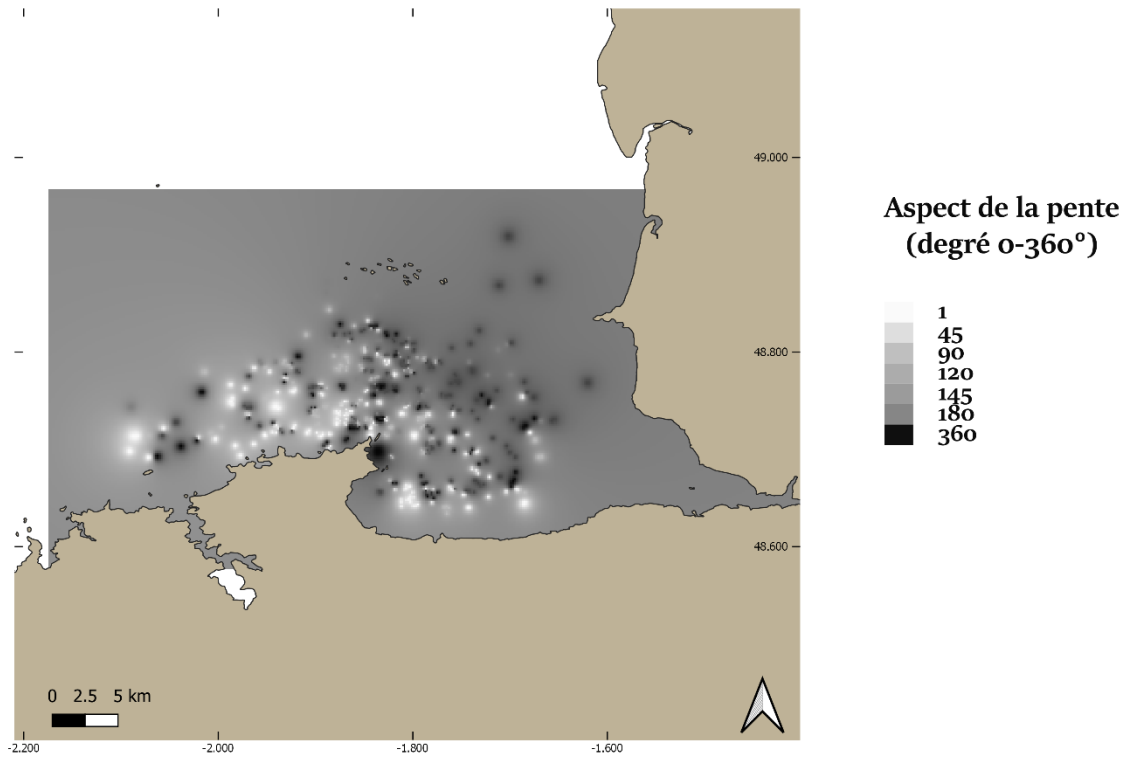


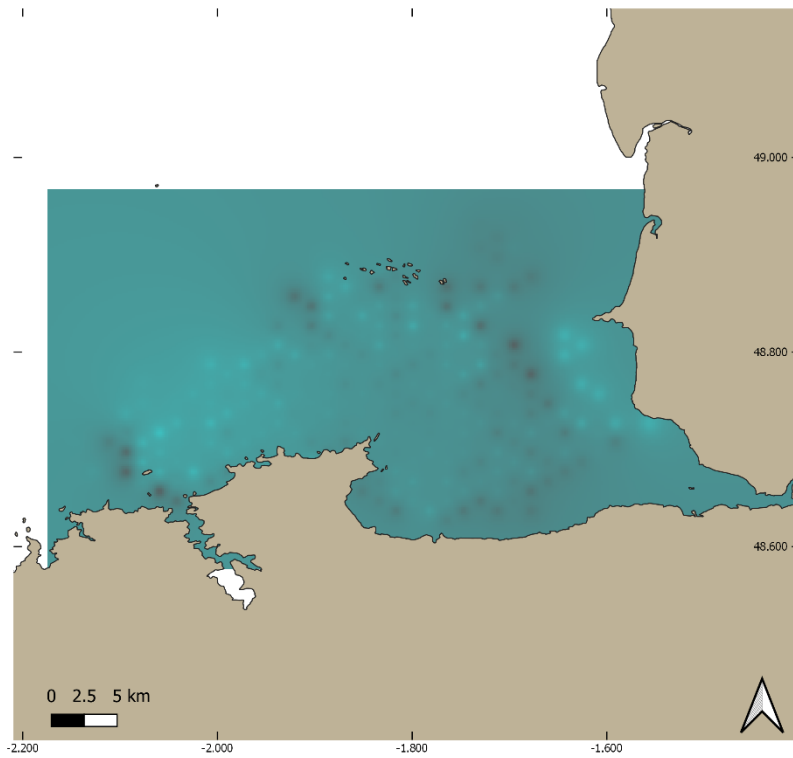
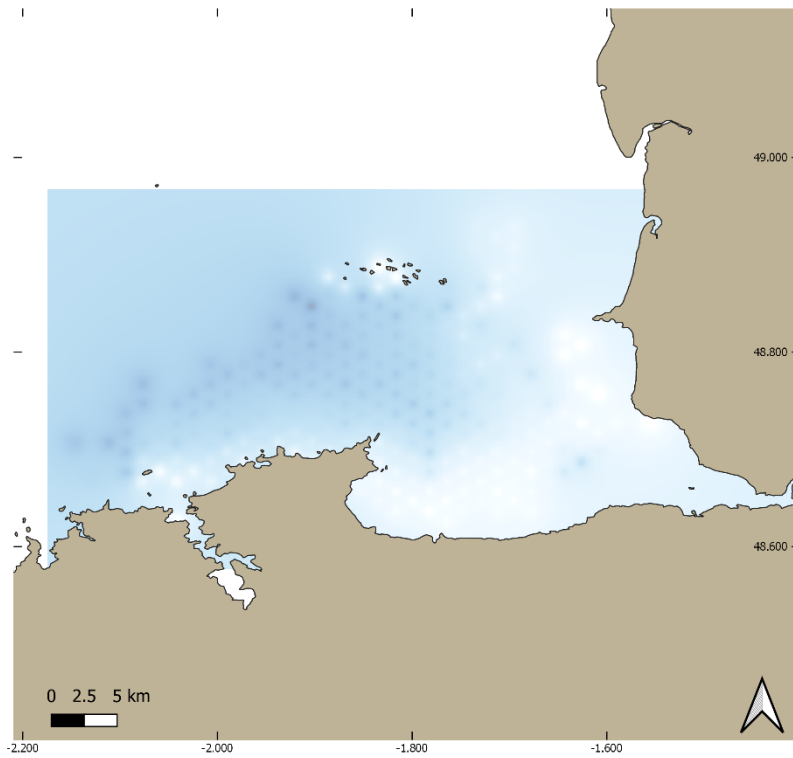
Fig.ci contre : Représentation du calcul de la distance géométrique moyenne (Hirzel 2005).

Calcule la moyenne géométrique des distances, le principe étant de dessiner le champ d'influence dans l'espace factoriel de chaque point d'observation de telle façon à ce que lorsqu'ils sont proches, leurs pouvoirs d'attraction se renforcent l'un l'autre. On calcule par point la moyenne géométrique des distances à chacun des points d'occurrence. Plus les points seront denses dans une zone, plus la probabilité que l'habitat est approprié pour l'espèce sera forte (Fig.). Aucune hypothèse n'est faite au préalable sur la forme de la distribution, mais la densité des observations doit être représentative de la niche de l'espèce. La moyenne géométrique produit un jeu d'enveloppes lissées autour des observations et une bonne généralisation de la niche.

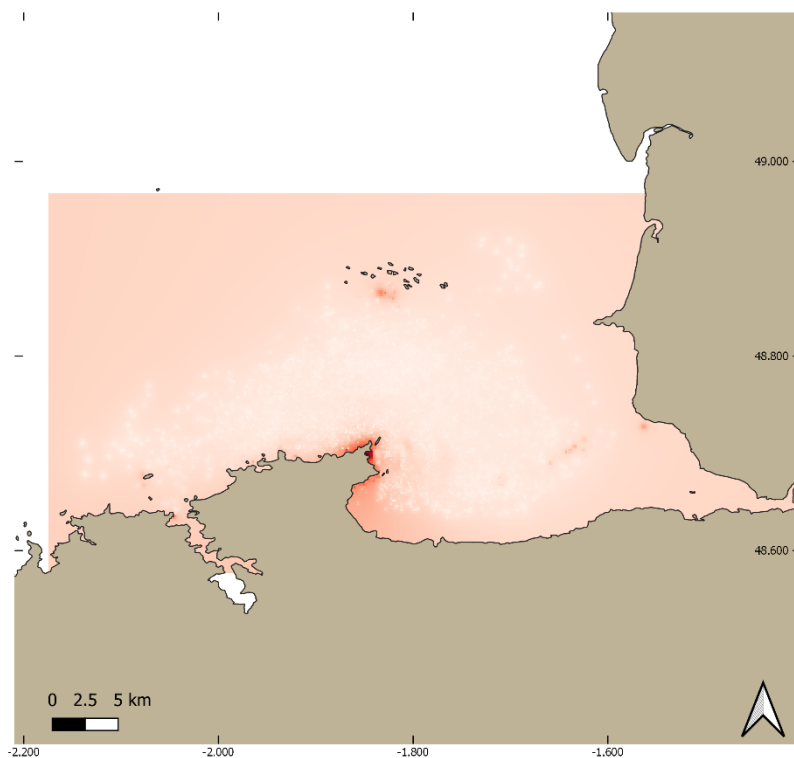


Annexe 13. Cartes éco-géographiques créées sur QGIS d'après le protocole décrit dans le présent rapport

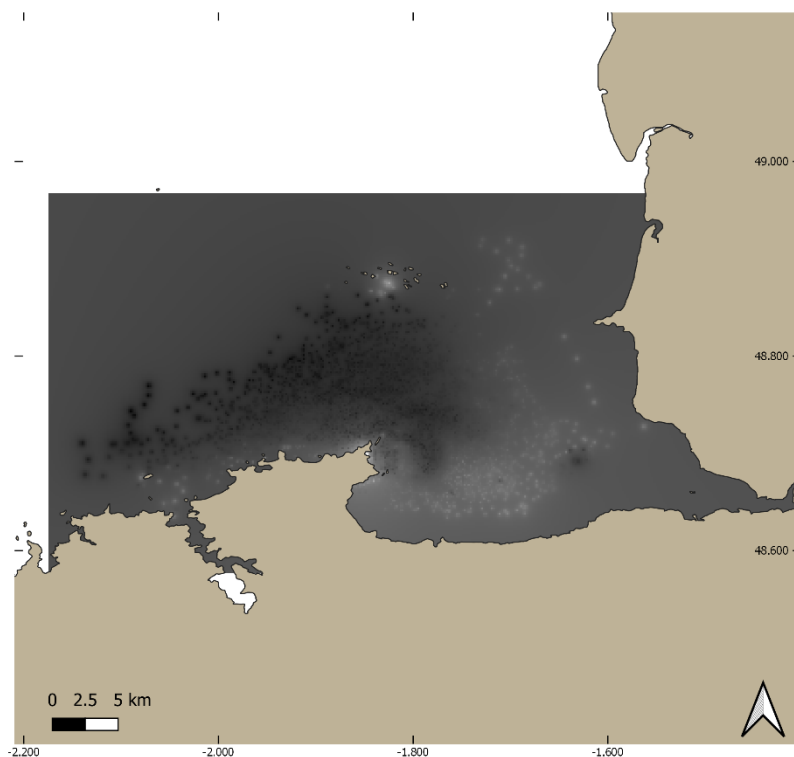
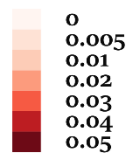




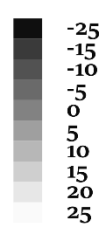


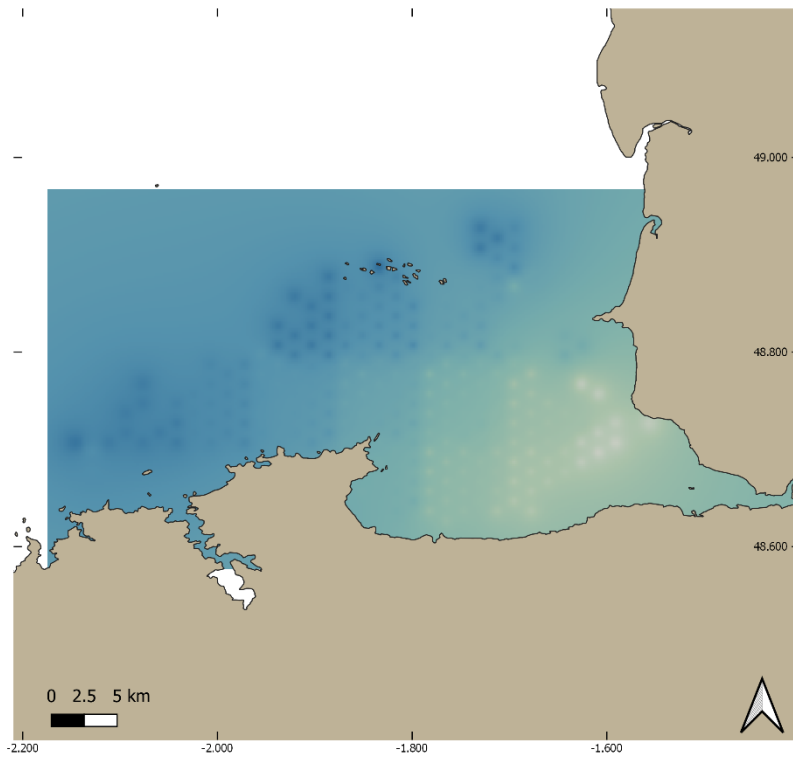


**Gradient de la pente  
(pourcentage d'élévation)**

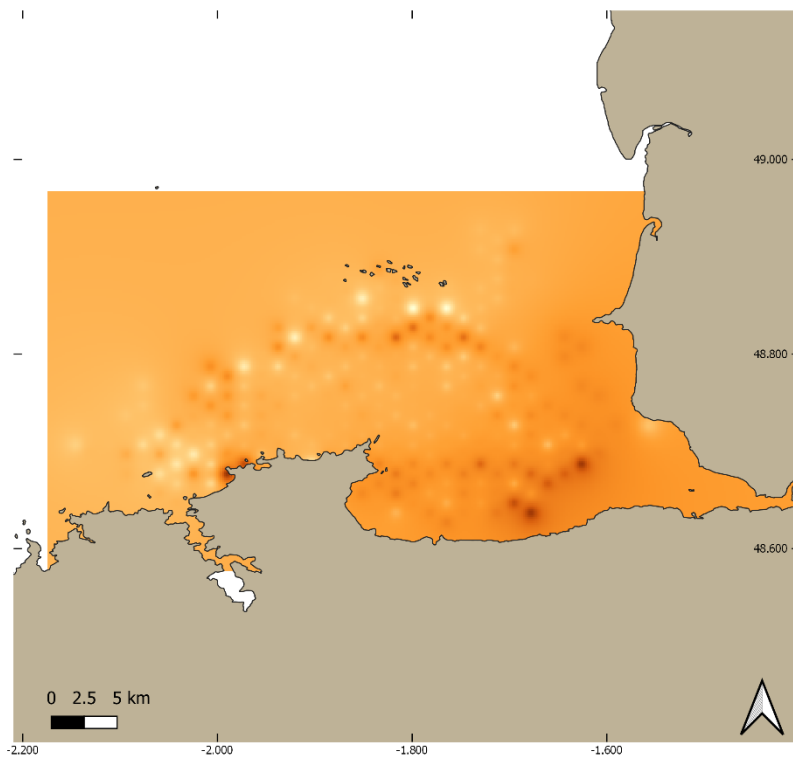


**Profondeur bathymétrique  
(m)**

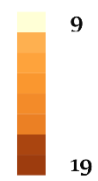


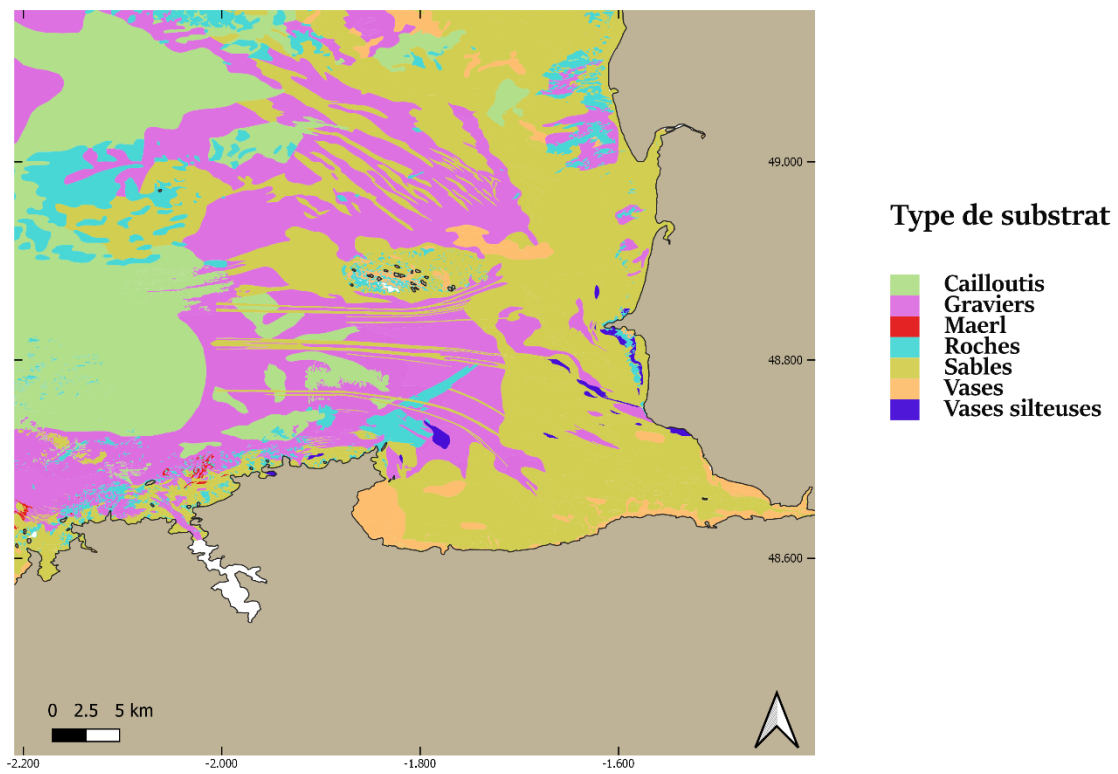
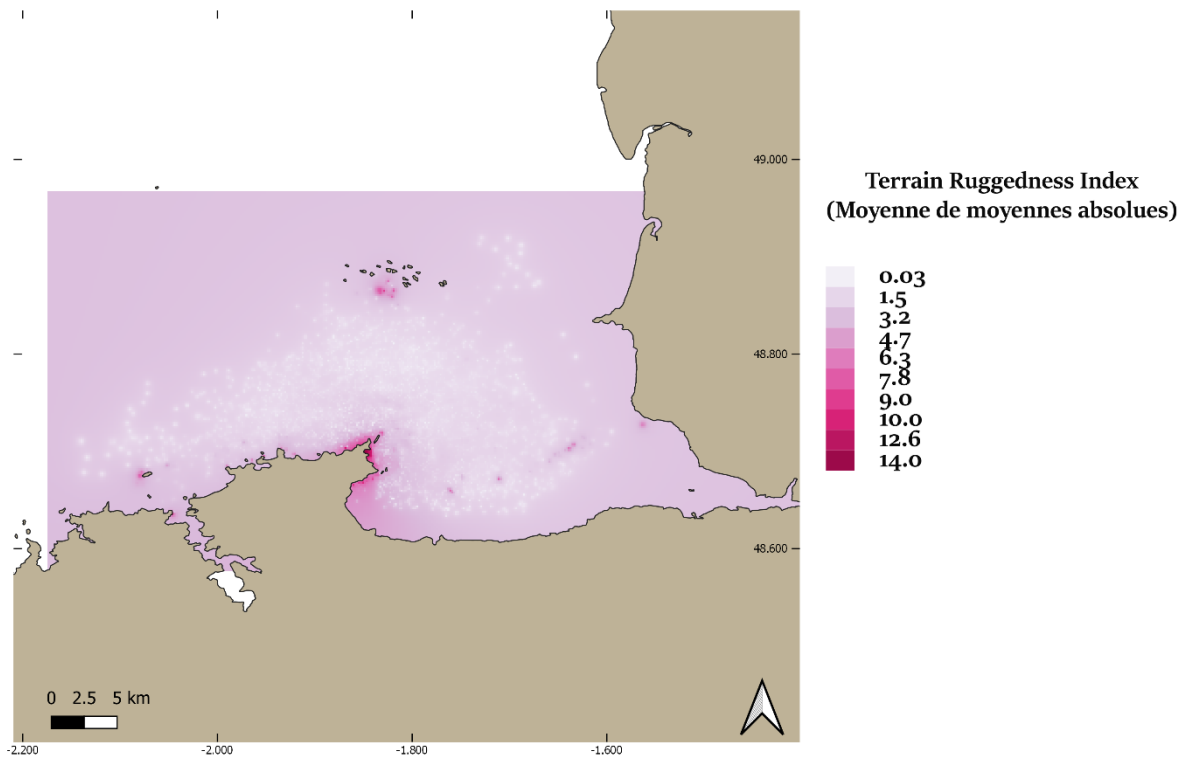


**Salinité  
(mg/m<sup>3</sup>)**



**Température de l'eau  
de surface (°C)**



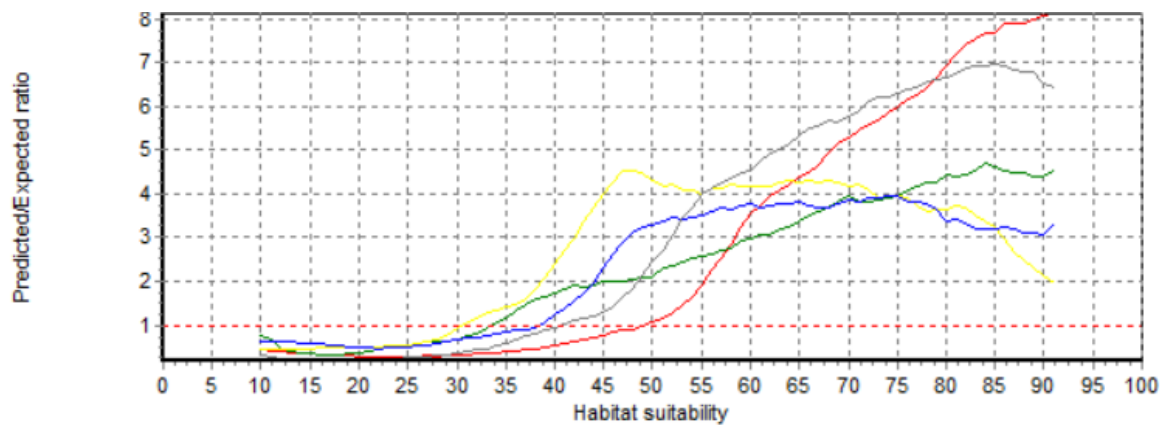


Annexe 14. Résultats de l'analyse ENFA à 4 variables

○ ENFA 4 variables

EGV's	Marginalité (19%)	SF1 (48%)	SF2 (21%)	SF3 (12%)
Chlorophylle	0.005	0.997	0.592	-0.441
Profondeur	0.324	-0.079	0.758	-0.299
Rugosité	-0.945	-0.021	0.259	-0.142
Distance	-0.043	-0.008	0.088	0.834
Marginalité:	0.607	Tolérance (1/S):		1.168
Spécialisation:	0.856	Indice de Boyce:		0.601 ± 0.55

Annexe 15. Courbes P/E à 6 variables 4 facteurs



Annexe 16. Tableau récapitulatif des valeurs d'AUC pour les modèles GLM et GAM à 6 ou 4 variables selon la méthode de génération d'absence utilisée (absence réelle vs. Pseudos absences) (2 répétitions random pour la création des pseudos absences)

Type de modèle		GAM			GLM		
		Absences réelles	Pseudos absences 1	Pseudos absences 2	Absences réelles	Pseudos absences 1	Pseudos absences 2
6 Variables	Run1	0.666	0.954	0.975	0.638	0.937	0.96
	Run2	0.675	0.97	0.968	0.63	0.956	0.967
	Run3	0.661	0.977	0.966	0.617	0.965	0.964
	Run4	0.678	0.972	0.958	0.637	0.961	0.937
	Run5	0.645	0.956	0.955	0.59	0.938	0.49
	Run6	0.699	0.969	0.971	0.658	0.951	0.965
	Run7	0.656	0.948	0.977	0.636	0.939	0.963
	Run8	0.665	0.962	0.96	0.63	0.956	0.951
	Run9	0.682	0.967	0.964	0.66	0.957	0.968
	Run10	0.73	0.953	0.967	0.629	0.941	0.961
	Mean	0.6757	0.96445		0.6325	0.93135	
Ecart-Type	0.02424894	0.008413492		0.019834034	0.104433446		
4 variables	Run1	0.662	0.951	0.957	0.646	0.937	0.947
	Run2	0.64	0.966	0.951	0.601	0.935	0.936
	Run3	0.654	0.968	0.962	0.616	0.942	0.948
	Run4	0.654	0.961	0.97	0.601	0.934	0.953
	Run5	0.624	0.969	0.961	0.606	0.947	0.955
	Run6	0.649	0.958	0.977	0.631	0.945	0.968
	Run7	0.668	0.952	0.967	0.628	0.933	0.946
	Run8	0.669	0.949	0.96	0.634	0.926	0.953
	Run9	0.685	0.963	0.963	0.617	0.938	0.951
	Run10	0.666	0.968	0.962	0.652	0.927	0.935
	Mean	0.6571	0.96175		0.6232	0.9428	
Ecart-Type	0.017045364	0.007290044		0.018017276	0.010430724		

Annexe 17. Importance des variables pour les modèles GAM à 6 ou 8 variables. 10 cross validation (Run) et deux répétitions de pseudos absences random (PA)

Var.env	GAM - 735 Présence, 735 Pseudos-absences- 6 variables																			
	PA1										PA2									
	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Chlorophylle-a	0.145	0.134	0.107	0.138	0.132	0.135	0.141	0.164	0.125	0.107	0.139	0.126	0.109	0.121	0.13	0.082	0.099	0.057	0.09	0.118
Distance à la côte	0.528	0.498	0.501	0.532	0.507	0.504	0.499	0.521	0.504	0.49	0.515	0.482	0.541	0.51	0.534	0.474	0.537	0.545	0.491	0.541
Profondeur	0.129	0.108	0.192	0.148	0.069	0.128	0.161	0.099	0.15	0.21	0.174	0.181	0.159	0.178	0.203	0.104	0.146	0.142	0.184	0.13
SST (°C)	0.057	0.035	0.044	0.073	0.083	0.037	0.045	0.033	0.056	0.059	0.047	0.039	0.031	0.05	0.083	0.039	0.031	0.056	0.067	0.024
Rugosité	0.634	0.0621	0.659	0.618	0.642	0.611	0.63	0.66	0.633	0.611	0.573	0.606	0.595	0.616	0.593	0.605	0.598	0.613	0.591	0.6
Tide Level	0.017	0.031	0.021	0.016	0.011	0.034	0.016	0.018	0.019	0.016	0.061	0.054	0.057	0.061	0.026	0.131	0.058	0.106	0.06	0.045

Var.env	GAM - 735 Présence, 735 Pseudos-absences- Modèle 4 variables																			
	PA1										PA2									
	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Chlorophylle-a	0.371	0.349	0.314	0.328	0.343	0.327	0.322	0.32	0.328	0.299	0.419	0.419	0.315	0.351	0.385	0.406	0.342	0.373	0.352	0.361
Distance à la côte	0.46	0.475	0.482	0.476	0.479	0.47	0.44	0.428	0.441	0.486	0.484	0.484	0.482	0.479	0.477	0.494	0.471	0.465	0.484	0.436
Profondeur	0.242	0.19	0.177	0.178	0.208	0.166	0.168	0.201	0.169	0.134	0.457	0.457	0.156	0.342	0.33	0.361	0.38	0.309	0.223	0.393
Rugosité	0.564	0.63	0.663	0.637	0.634	0.639	0.664	0.603	0.653	0.651	0.58	0.58	0.635	0.6	0.596	0.59	0.575	0.586	0.609	0.608

Annexe 18. Importance des variables pour les modèles GLM à 6 ou 8 variables. 10 cross validation (Run) et deux répétitions de pseudos absences random (PA)

Var.env	GLM - 735 Présence, 735 Pseudos-absences- Modèle 6 variables																			
	PA1										PA2									
	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Chlorophylle-a	0.089	0.05	0.02	0.04	0.03	0.08	0.052	0.03	0.06	0.051	0.06	0.05	0.04	0.07	0.02	0.06	0.05	0.03	0.06	0.055
Distance à la côte	0.512	0.05	0.49	0.5	0.52	0.51	0.49	0.522	0.506	0.495	0.52	0.54	0.53	0.49	0.54	0.49	0.53	0.53	0.49	0.552
Profondeur	0.131	0.19	0.19	0.21	0.1	0.17	0.175	0.153	0.206	0.157	0.31	0.26	0.21	0.28	0.24	0.25	0.22	0.22	0.3	0.264
SST (°C)	0.048	0.08	0.11	0.09	0.07	0.04	0.07	0.075	0.064	0.054	0.12	0.08	0.06	0.09	0.1	0.09	0.07	0.09	0.1	0.066
Rugosité	0.656	0.68	0.7	0.66	0.69	0.65	0.675	0.688	0.675	0.674	0.62	0.66	0.66	0.67	0.66	0.65	0.67	0.69	0.68	0.642
Tide Level	0	0	0	0	0	0.03	0	0.029	0.027	0.023	0	0.04	0.04	0	0.03	0	0.05	0.04	0	0.042

Var.env	GLM - 735 Présence, 735 Pseudos-absences- Modèle 4 variables																			
	PA1										PA2									
	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
Chlorophylle-a	0.218	0.18	0.16	0.16	0.18	0.19	0.177	0.164	0.176	0.169	0.2	0.18	0.15	0.25	0.21	0.23	0.19	0.17	0.2	0.175
Distance à la côte	0.467	0.48	0.48	0.48	0.47	0.47	0.452	0.451	0.547	0.489	0.45	0.4	0.47	0.44	0.43	0.45	0.46	0.44	0.44	0.456
Profondeur	0.175	0.14	0.16	0.19	0.17	0.16	0.152	0.186	0.187	0.179	0.27	0.24	0.16	0.26	0.23	0.27	0.22	0.21	0.19	0.263
Rugosité	0.608	0.62	0.65	0.65	0.61	0.36	0.631	0.635	0.676	0.637	0.59	0.67	0.66	0.55	0.62	0.56	0.54	0.62	0.62	0.12

Annexe 19. Importance des variables pour les GLM et les GAM individuels

	PA1										PA2										MEAN		
GLM individuels	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Fullmodel1	Fullmodel2	
rasterchloro	0.175	0.093	0.124	0.115	0.174	0.17	0.138	0.128	0.171	0.11	0.128	0.186	0.17	0.136	0.135	0.156	0.142	0.121	0.128	0.143	0.175	0.146	0.14381818
rasterdistance	0.466	0.476	0.495	0.463	0.476	0.486	0.471	0.493	0.492	0.472	0.467	0.492	0.478	0.501	0.483	0.478	0.48	0.486	0.509	0.48	0.499	0.484	0.48304545
rasterProfondeurmean	0.175	0.111	0.111	0.137	0.145	0.137	0.121	0.131	0.127	0.122	0.128	0.159	0.243	0.202	0.137	0.21	0.179	0.174	0.148	0.208	0.194	0.165	0.15745455
persistrasterroughness	0.631	0.702	0.657	0.661	0.617	0.62	0.639	0.65	0.633	0.658	0.647	0.577	0.603	0.618	0.628	0.615	0.632	0.637	0.589	0.627	0.606	0.609	0.62981818

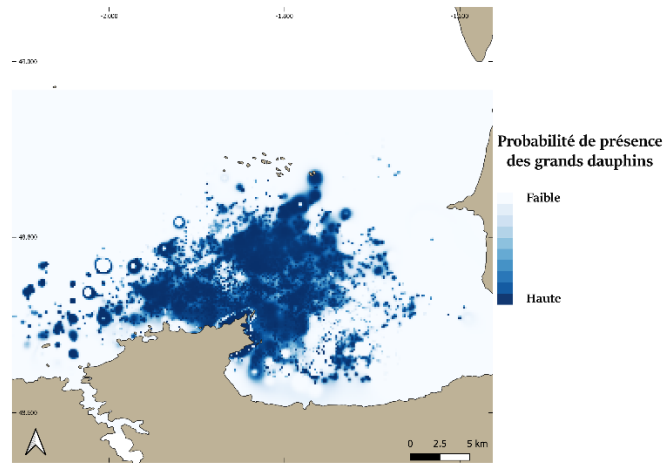
  

	PA1										PA2										MEAN		
GAM individuels	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10	Fullmodel1	Fullmodel2	
rasterchloro	0.297	0.34	0.309	0.33	0.346	0.345	0.299	0.349	0.359	0.339	0.338	0.405	0.352	0.396	0.42	0.331	0.39	0.394	0.363	0.358	0.343	0.365	0.35309091
rasterdistance	0.463	0.441	0.465	0.422	0.43	0.393	0.456	0.473	0.472	0.409	0.434	0.415	0.44	0.446	0.438	0.439	0.44	0.423	0.43	0.494	0.434	0.449	0.44118182
rasterProfondeurmean	0.08	0.257	0.17	0.178	0.246	0.256	0.139	0.219	0.214	0.218	0.211	0.43	0.306	0.343	0.402	0.271	0.346	0.356	0.301	0.287	0.27	0.321	0.26459091
persistrasterroughness	0.651	0.66	0.605	0.661	0.617	0.634	0.684	0.596	0.595	0.647	0.641	0.588	0.617	0.63	0.575	0.592	0.606	0.61	0.624	0.624	0.655	0.616	0.624

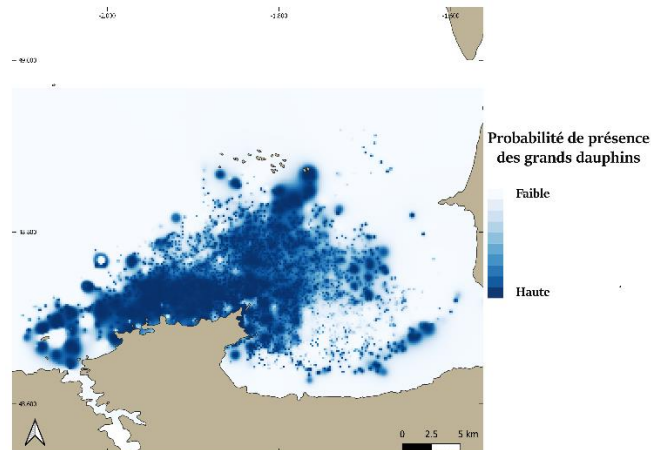
Annexe 20. Valeurs d'AUC pour les modèles GAM et GLM individuels, et pour les modèles GAM et GLM d'ensemble

GLM	0.966	0.928	0.945	0.95	0.949	0.951	0.949	0.953	0.957	0.939	0.953	0.939	0.943	0.951	0.935	0.949	0.932	0.959	0.946	0.957	0.941	0.949
GLM Modèle d'ensemble	0.951																					
GAM	0.965	0.973	0.964	0.972	0.98	0.973	0.933	0.979	0.961	0.971	0.973	0.947	0.958	0.945	0.933	0.945	0.94	0.961	0.959	0.954	0.958	0.964
GAM Modèle d'ensemble	0.968																					

Annexe 21. Carte de probabilité de présence des grands dauphins selon le meilleur modèle individuel GAM



Annexe 22. Carte de probabilité de présence des grands dauphins selon le meilleur modèle individuel GLM



## Annexe 23. Fiche projet OCEAN HACKATHON 2020

Titre

### Morgazh: l'outil d'extraction de données environnementales

#### Description

Depuis 2004, l'association AL LARK collecte des données sur la population de grands dauphins du golfe normano-breton: données GPS, environnementales, comportementales, spatio-temporelles. Le but est de développer les connaissances sur cette population dans un objectif de conservation.

Nous souhaiterions intégrer des données environnementales supplémentaires non collectées par AL LARK (température, salinité, chlorophylle, données halieutiques, etc.) afin de caractériser de manière plus précise les conditions de distribution des dauphins.

L'ensemble de ces données provenant de nombreuses sources et se présentant sous des formats très variables, nous souhaiterions créer un outil permettant de les centraliser sur une base de données, en fonction des coordonnées GPS et de la date et heure de nos données. L'outil devra être capable d'aller récupérer ces données supplémentaires sur des sources externes (Ex: Copernicus pour la chlorophylle, le SHOM pour la bathymétrie).

Il devra regrouper les données sous un format csv afin de permettre son exploitation statistique et cartographique.

#### Concrètement, que voudriez-vous développer en 48h ?

Une interface facile d'utilisation, y compris par une personne ne sachant pas coder. Un outil qui, sur une période donnée:

- va chercher les données environnementales choisies à la source,
- les centralise sous un même format,
- met en relation ces données avec nos données collectées, selon les coordonnées GPS, date et heure,
- extrait dans un nouveau fichier, sous un format csv, toutes ces données aux points GPS, date et heure demandés,
- visualisation des erreurs d'extraction (Ex: données manquantes).

#### Définissez 4 mots clés qualifiant votre défi

Développement informatique / BDD ; SIG/rasters ; codage ; programmation web.

#### Pourriez-vous décrire brièvement ces données

Données binaires, qualitatives, quantitatives, horaire, date, coordonnées spatiales.

#### De quelles données auriez-vous besoin pour réaliser votre défi ?

Des données environnementales qui ne sont pas collectées à bord, mais accessibles via des sources externes: température, salinité, chlorophylle A, données halieutiques, hauteur d'eau, marée, courantologie, nitrate, micropolluants, etc.

#### Je recherche des équipiers spécialistes de :

Développement informatique / BDD ; SIG/cartographie ; codage informatique.

#### Sur quelle zone géographique s'applique plus particulièrement votre défi?

Golfe Normano-breton

#### Vous proposez votre défi en tant que ? Structure morale, pour l'association AL LARK

## Résumé

La modélisation de la distribution des dauphins est essentielle pour comprendre leur écologie et pour optimiser leur conservation et leur gestion. Les informations sur la distribution et les habitats préférentiels des Grands Dauphins (*Tursiops truncatus*) du golfe normand-breton sont peu nombreuses, en particulier dans la zone de la Baie du Mont Saint Michel où l'espèce peut être menacée par de nombreuses activités anthropiques. Ici, l'utilisation de données d'observations récoltées lors de sorties en mer est couplée à une approche de modélisation afin d'identifier les zones où la probabilité d'occurrence du grand dauphin est élevée ainsi que les variables environnementales influant sur leur distribution. Quatre techniques de modélisation sont utilisées afin de pouvoir comparer leur pouvoir de prédiction et de déterminer quel type de modèle est le plus adapté aux données récoltées par l'association. L'analyse des facteurs de niche écologique (ENFA) utilise des données de présence seulement, tandis que les modèles linéaires généralisés (GLM) et les modèles additifs généralisés (GAM) utilisent des données de présence/absence. Une comparaison de ces modèles avec une méthode de modélisation d'ensemble sera effectuée afin de déterminer si cette approche est plus performante que les modèles individuels. Les données ont été collectées sur quatre ans (de mars 2016 à novembre 2019) et représentent 741 sorties et 494 observations de grands dauphins. L'approche de modélisation de l'ENFA semble plus adaptée aux données disponibles pour cette étude, tandis que les modèles d'ensemble ne sont pas toujours plus performants que les GLM et les GAM ici. Les résultats montrent que la rugosité du fond, la profondeur, et la distance à la côte sont les variables environnementales influant sur la distribution de l'espèce. De plus, cette étude révèle les habitats les plus adaptés pour les grands dauphins de la zone d'étude, ce qui permet d'apporter une compréhension plus précise et complète de la répartition de la population, de définir de nouvelles zones de prospection et de souligner l'intérêt des modèles de distribution dans l'élaboration de stratégies efficaces de conservation pour cette espèce.

### Mots clés :

Cétacés – Modèles de distribution d'espèce - Analyse des facteurs de niche écologique – *Tursiops truncatus* – Ecologie marine – Golfe normand-breton



## Abstract

Modelling dolphin distribution is key for understanding their ecology and for their conservation and management. Information on the distribution and preferred habitats of Normano-Breton Gulf bottlenose dolphins (*Tursiops* sp.) is lacking, particularly in the Mont Saint Michel Bay, where the species could be threatened by several anthropogenic activities.

Here, the use of boat survey sampling data sampled is coupled with a modeling approach to identify areas of high probability of bottlenose dolphin occurrence and the environmental variables influencing their distribution. Four modeling techniques are used in order to compare their predictive power and to determine which type of model is the most adapted to the data collected by the association. The Ecological Niche Factor Analysis (ENFA) uses presence data only, while Generalized Linear Models (GLM) and Generalized Additive Models (GAM) use presence/absence data. A comparison of these models with an aggregate modeling approach will be conducted to determine if this approach is more efficient than individual models.

The data were collected over four years (March 2016 to November 2019), and represent 741 trips and 494 sightings of bottlenose dolphins. The ENFA modeling approach appears to be more appropriate to the data available for this study, while ensemble models don't always perform better than GLM and GAM in all cases. The results show that bottom roughness, depth, and distance from the shore are the environmental variables influencing the distribution of the species. In addition, this study reveals the most suitable habitats for bottlenose dolphins, providing a more accurate and complete understanding of the distribution of the species in the area, defining new areas for prospecting and highlighting the interest of distribution models in the development of effective conservation strategies for this species.

## Keywords :

Cetaceans - Species distribution models - Environmental niche factor analysis – *Tursiops truncatus*  
– Marine ecology – Normano-Breton Gulf

## Resumen

La elaboración de modelos de distribución de los delfines es esencial para entender su ecología y para su conservación y gestión. La información sobre la distribución y los hábitats preferidos de los delfines mulares (*Tursiops truncatus*) en el Normano-Breton Golfo es escasa, especialmente en la zona de la bahía del Mont Saint Michel, donde la especie puede estar amenazada por numerosas actividades antropogénicas. En este caso, el uso de los datos de observación recogidos durante los viajes por mar se combina con un enfoque de modelización a fin de identificar las zonas en las que la probabilidad de que se produzcan delfines mulares es alta y las variables ambientales que influyen en su distribución.

Se utilizan cuatro técnicas de modelización para comparar su poder de predicción y determinar qué tipo de modelo es el más adaptado a los datos recogidos por la asociación. El análisis del factor de nicho ecológico (ENFA) utiliza únicamente datos de presencia, mientras que los modelos lineales generalizados (GLM) y los modelos aditivos generalizados (GAM) utilizan datos de presencia/ausencia. Se realizará una comparación de estos modelos con un enfoque de modelización global para determinar si este enfoque es más eficiente que los modelos individuales.

Los datos se recopilaron durante cuatro años (de marzo de 2016 a noviembre de 2019) y representan 741 viajes y 494 avistamientos de delfines mulares. El enfoque de modelización del ENFA parece más adaptado a los datos disponibles para este estudio, mientras que los modelos de conjunto no siempre funcionan mejor que el GLM y el GAM en algunos casos. Los resultados muestran que la rugosidad del fondo, la profundidad y la distancia de la costa son las variables ambientales que influyen en la distribución de la especie. Además, este estudio revela los hábitats más adaptados para los delfines mulares, lo que permite comprender de manera más precisa y completa la distribución de la especie en la zona, definir nuevas zonas de prospección y subrayar el interés de los modelos de distribución en la elaboración de estrategias eficaces de conservación de esta especie.

## Palabras clave :

Cetáceos - Modelos de distribución de especies - Análisis de los factores de nichos ecológicos - *Tursiops truncatus* - Ecología marina – Normano-Breton Golfo - Modelos de distribución de especies