

الجمهورية الجزائرية الديمقراطية الشعبية

République algérienne démocratique et populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement supérieur et de la recherche scientifique

المدرسة الوطنية العليا لعلوم البحر وتهيئة الساحل

École nationale supérieure des sciences de la mer et de l'aménagement du littoral



Mémoire de fin d'étude en vue de l'obtention du diplôme master 2

En l'environnement marin et côtier

Option : ingénierie en environnement marin et côtier

Thème :

*Application de l'intelligence artificielle pour la modélisation
spatio-temporelle du grand pélagique (Thunnus thynnus)
en relation avec les facteurs environnementaux dans le
bassin Méditerranéen*

Présenté par :

Beldjoudi Sara & Bousboula Hadil

Encadré par :

M. Benyahia Firad & M. Boufeniza redouane larbi

Soutenue publiquement le : 28/06/2026,

Devant le jury composé :

M. FIRAD.B	MCB Promoteur	(ENSSMAL)
M. BOUFENIZA.RL	Chercheur Co-promoteur	(NUIST)
Mme. Ait Aissa Djamila	MCA présidente	(ENSSMAL)
M. Khames Yacine	MCA examinateur	(USTHB)

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Remerciements

Avant tout, nous remercions Allah, le Tout-Puissant, de nous avoir donné la force, le courage nécessaire pour mener à bien ce travail.

Nous tenons à exprimer notre profonde gratitude à notre promoteur. Dr. Benyahia Firad, pour son accompagnement, sa disponibilité, ses précieux conseils et son soutien tout au long de la réalisation de ce projet. Ses orientations ont été d'une grande aide pour mener ce travail à terme. Merci infiniment

Nos remerciements s'adressent également à notre Co-promoteur, Dr. Boufeniza Redouane larbi pour son accompagnement, ses conseils avisés, sa disponibilité et son soutien tout au long de la réalisation de ce travail.

Nous adressons également nos sincères remerciements aux membres du jury à Mme.Ait Aissa Djamila, président du jury,

Nous remercions vivement M. Khames Yacine D'avoir accepté d'examiner ce travail et pour l'intérêt qu'il lui a porté.

Nous exprimons aussi notre reconnaissance à tous nos enseignants de L'ENSSMAL, qui ont contribué à notre formation.

Enfin, nous remercions toutes les personnes qui ont contribué, de près ou de loin, direct ou indirect à la réalisation de ce travail.

Dédicace

This project is for those who have shaped my path, inspired my growth, and filled my life with love and laughter.

To my amazing, partner Hadil, thank you for standing with me through every challenge and success. Our partnership turned what could have been a lonely journey into something I will always cherish. This work is not just mine; every page carries both our signatures, etched in fatigue, laughter, and persistence.

To my beloved parents, you've been the silent fuel behind all my efforts. Your sacrifices never made headlines, but they shaped my world—small kindnesses, steady encouragement, and that quiet courage when things got tough. When I move forward, it's your hopes I carry, your prayers smoothing the path beneath my feet. No words ever feel enough, but I hope you see in me a reflection of everything you've poured into my life.

To my brothes, thank you for your support and encouragement throughout this journey. A special mention to my brother Zaki. Thank you for always encouraging me and believing in my abilities. I truly appreciate your support.

To my sister Ahlam, Blood can't explain our bond, because what we have goes deeper than genetics. You are, in every sense, my sister—the one whose advice keeps me steady and whose warmth fills my empty spaces. You've celebrated my wins, offered a hand through my messes, and filled my life with understanding. Thank you for being my haven and always standing in my corner.

To my nieces, Rana and Djana, I see so much hope in your bright eyes. Your laughter pulls me out of the hardest days, and your hugs remind me what I'm really working for. When you look at this diploma or hear about my journey, I want you to know that dreams don't have limits if you carry them with love and courage. Aunt Sara is beyond proud to have you growing alongside me.

To my dear roommate, best friend and soul sister Lina, we crammed five years of life into those four walls making a home out of thin air, we built our own world inside those four walls, it's hard to find words for what you've become to me: family, anchor, daily reminder that home is who, not where. the unforgettable years we shared are treasures I'll always Keep.

You made each day brighter with your laughter and support. Graduation may lead us to different paths but our memories and bond will stay.

The thought of saying goodbye sits heavy on my chest. I can't imagine waking up and not finding you there. You turned routine into magic and shaped my university years into the happiest memories I'll ever own. Though we're starting new chapters apart, your laughter and love echo in every corner of my heart.

This project belongs to each of you. You pushed me, held me, believed in me before I could believe in myself, and made every step possible. To the people who gave me unwavering love, trust, and hope—this is for you.

All my love and endless gratitude,

Sara

Dédicace

*Je commence par remercier Allah,
Celui qui m'a portée dans mes moments de faiblesse, éclairée dans mes doutes, et donné la force de continuer quand tout me semblait difficile. Sans Sa miséricorde, je ne serais jamais arrivée jusqu'ici.*

À ma mère, À toi... les mots ne suffiront jamais. Tu es bien plus qu'une mère, tu es mon âme, ma lumière, mon refuge. Tu as cru en moi quand je doutais, tu as essuyé mes larmes sans jamais te plaindre, tu as porté mes peines en silence pour me voir sourire. Chaque réussite que j'accomplis porte ton nom. Je t'aime d'un amour que rien ne peut mesurer.

À mon père, Toi qui as été mon pilier dans chaque étape de ma vie, mon exemple de courage et de dignité. Tu m'as appris à ne jamais abandonner, à me relever après chaque chute. Merci d'avoir toujours cru en moi, même quand je ne croyais plus en moi-même. Ce succès est le tien avant d'être le mien.

À ma seule sœur, Ma confidente, mon soutien silencieux, ma force au quotidien. Merci d'avoir été là sans condition, dans les bons comme dans les mauvais moments. Ta présence est un cadeau que la vie m'a offert.

À moi-même, Pour toutes ces nuits de doute, ces moments de fatigue, ces larmes cachées... et pourtant, je ne me suis pas arrêtée. Je suis restée debout, j'ai avancé, j'ai résisté. Aujourd'hui, je suis, fière de ce chemin rempli d'efforts et de courage.

À ma famille, Vous êtes mon ancrage, ma force et mon équilibre dans cette vie. Merci pour votre amour, votre présence et votre soutien constant qui m'ont accompagnée à chaque étape de ce parcours. Une pensée toute particulière à mon cher khali Azzouz, ainsi qu'à Salim et Fateh, pour leur bienveillance et leur encouragement. À toutes mes tantes, Noura, Nassira et Radia, merci pour votre affection et votre attention qui m'ont profondément touchée. À mes chers cousins, Bassim et Kelthoum, qui ont toujours pris de mes nouvelles avec tant de gentillesse, Et surtout, à mon précieux Jedo, dont la présence est une bénédiction dans ma vie. Sans oublier mes adorables bébés Yazane et Zozaa, loulou et firas qui, par leur

Innocence et leur joie, ont illuminé mes moments les plus difficiles. Vous êtes tous une partie de mon cœur,

À ma meilleure binôme Sara, Plus qu'une simple partenaire, tu as été une véritable alliée dans cette aventure. Nous avons partagé le stress, les difficultés, mais aussi les moments de joie. Merci pour ta patience, ton engagement et ta sincérité. Je garderai de cette expérience des souvenirs inoubliables.

À ma chère Meryem, Merci pour tes mots qui remontent le moral dans les moments difficiles. Merci pour tous les petits moments partagés, les rires, le soutien et ta présence précieuse. Je suis reconnaissante de t'avoir dans ma vie.

À ma précieuse amie Joujou, Tu as été cette lumière dans mes moments sombres, ce sourire dans mes journées difficiles. Merci d'avoir toujours été là, sans jamais faillir. Ta présence a été un véritable soutien pour moi, et ta place dans mon cœur est unique et éternelle. Enfin, À toutes les personnes qui ont contribué, de près ou de loin, à ce travail... merci du fond du cœur.

Avec tout mon amour, ma reconnaissance et une émotion sincère.

Hadil

Table des matières

Liste des figures	9
Liste des Tableaux	11
Liste des Abréviations.....	12
Liste des Symboles.....	14
Résumé :.....	15
Abstract:	16
: ملخص.....	17
Introduction	18
Chapitre I : Synthèse Bibliographique et Généralité	22
1. Recherche bibliographique	23
1.1 Stratégie de recherche documentaire	23
1.2. Analyse du réseau institutionnel de la recherche	25
1.3. Synthèse basée sur les modèles d'IA	25
1.4. Synthèse des résultats bibliométriques et méthodologiques	26
2. Le grand pélagique (<i>Thunnus thynnus</i>).....	27
2.1. Importance écologique, économique et pression anthropique	27
3. Biologie et écologie du thon rouge (<i>Thunnus thynnus</i>).....	27
3.1. Taxonomie et adaptations physiologiques	27
3.2 Morphologie	28
3.3 Reproduction migrations et utilisation de l'habitat en Méditerranée	29
4. Facteurs environnementaux contrôlant la distribution.....	30
4.1 Variables physiques	30
4.2 Variables biogéochimiques	31
4.3 Données mobilisables pour l'étude des pélagiques	32
4. 4 Données environnementales et océanographiques (télédétection, SST, chlorophylle)	32

Table des matières

4.5. Données biologiques et halieutiques (captures, campagnes scientifiques, marquage)	32
5. Modélisation de la distribution, intelligence artificielle et données	32
5.1 Modèles de distribution des espèces : principes et limites des approches classiques	32
5.2 Apprentissage automatique	33
5.3 Données mobilisables : télédétection, pêcheries et intégration multi-sources	33
6. Positionnement du projet	33
6.1. Lacunes scientifiques identifiées	33
6.2. Apports attendus de l'étude	34
Chapitre II : Matériels et Méthodes	35
1. Zone d'étude	36
1.1. Présentation générale du bassin méditerranéen	36
1.2 Base de données	38
1.3 Description des variables	40
1.4 Outils de programmation et logiciel	41
1.5 Description des algorithmes utilisés	42
2. Méthodologie	44
2.1 Le prétraitement des données	44
3. L'analyse statistique descriptive	47
3.1 Moyenne et médiane	47
3.2 Écart-type et variance	48
3.3 Valeurs minimales et maximales	48
3.4 Quartiles (Q1, Q2, Q3)	48
3.5 Asymétrie (Skewness)	48
3.6 Aplatissement	48
3.7 Matrice de corrélation	48
3.8 Histogrammes et Boxplots	48

Table des matières

4. Métriques et évaluation.....	49
5. Les différentes méthodes de fine-tuning.....	52
5.1 Optimisation des hyperparamètres	52
Chapitre III :	53
Résultats Et Discussions	53
1. Les Résultats	54
1.1. Prédiction de position de distribution de thon rouge (Target 1)	54
1.2. Prédiction de la Longueur du Thon Rouge (Targte 2)	60
1.3. Analyse Globale et Importance des Variables	64
1.4 Résultat de la corrélation observe vs prédit	67
2. Discussion Générale.....	86
Conclusion	91
Bibliographies	73
Annexes	79

Liste des figures

Figure 1: Nombre d'études utilisant différents modèles d'IA dans la région méditerranéenne et dans les océans Atlantique, Pacifique et Indien.....	26
Figure 2: Thon rouge de l'Atlantique (Thunnus thynnus)(NOAA, 2026).	29
Figure 3: Répartition spatiale, zones de reproduction (en vert) et principales voies migratoires du thon rouge (Thunnus thynnus) entre la Méditerranée et l'Atlantique Est(Eriksen et al., 2025)	30
Figure 4: Zone d'étude (bassin méditerranéen).	37
Figure 5: Schéma conceptuel des données utilisées pour la modélisation de Thunnus thynnus en Méditerranée.	40
Figure 6: Schéma synthétique du pipeline Prétraitement des données (Thunnusthynnus), Méditerranée Occidentale (2000–2019).	47
Figure 7: comparative des métriques (R^2 , RMSE, MAE) pour les quatre modèles, avant optimisation et après optimisation par recherche bayésienne et Grid Search.....	55
Figure 8: Densité spatiale du Thon Rouge (KDE — Ordinary Kriging) — Test set (Bassin Occidental).....	57
Figure 9: Erreur du modèle (Ordinary Kriging) — Test set (Bassin Occidental) Pixels 0.045° Palette blanc-noir XGBoost (target 1)	58
Figure 10: Histogramme des distances d'erreur géographique (individus/pixel) du modèle XGBoost (target 1) avec indication des métriques MAE, médiane et RMSE	59
Figure 11: Comparative des métriques (R^2 , RMSE, MAE) pour les quatre modèles, avant optimisation et après optimisation par recherche bayésienne de target 2 (Grid Search pour la régression linéaire).....	61
Figure 12: Distribution spatiale des classes de taille du thon rouge (petite, moyenne, grande taille) en Méditerranée occidentale — Test set — Random Forest Bayesian (target 2).....	62
Figure 13: Distribution spatiale de la magnitude des résidus (krigeage) — Test set Méditerranée occidentale — Random Forest Bayesian (target 2) — MAE = 21.94 cm, RMSE = 31.76 cm	63
Figure 14: Histogramme des résidus signés — Test set Méditerranée occidentale — Random Forest Bayesian (target 2)	64
Figure 15: Importance des variables pour target 1 (position LonX + LatY) — Modèle XGBoost + MultiOutputRegressor + Grid Search — Méditerranée	65

Liste des figures

Figure 16: Importance des variables (MDI et permutation) pour target 2 (LenCM) — Modèle Random Forest Bayesian— Méditerranée	66
Figure 17: corrélation spatiale de la densité observée et la densité prédite par le modèle xg boost dans le bassin Méditerranée — Densité observée (2000–2019) vs Densité prédite (2020–2025)	68
Figure 18: corrélation saisonnière de la densité spatiale du thon rouge dans le bassin Méditerranée — Densité observée vs Densité prédite par le modèle xg boost	69
Figure 19: corrélation annuelle de la densité spatiaux temporelle du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle XGBoost	71
Figure 20: corrélation mensuelle mensuelle de la densité spatiale prédite vs observée du thon rouge dans le bassin méditerranéen — par le modèle XGBoost	74
Figure 21: corrélation de la taille moyenne du thon rouge dans le bassin Méditerranée — Observé 2000-2019 vs Prédit 2020-2025 par le modèle random forest	75
Figure 22: corrélation mensuelle de la taille du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest	78
Figure 23: corrélation saisonnière de la taille du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest	79
Figure 24: corrélation annuelle de la taille du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest	81
Figure 25: Taille moyenne et classes biologiques du thon rouge (<i>Thunnus thynnus</i>) en Méditerranée — corrélation entre les observations (2000-2019) et les prédictions (2020-2025)	83
Figure 26: Répartition des classes de taille (petite, moyenne, grande) : effectifs et proportions relatives	84
Figure 27: Écart (en %) entre la prédiction du modèle et l'observation – par classe de taille	85

Liste des Tableaux

Tableau 1: Étapes des critères de sélection des données.....	23
Tableau 2: Répartition du domaine d'étude des 27 articles sélectionnés.....	24
Tableau 3: Classement des institutions de recherche méditerranéennes par citations et thématique.....	25
Tableau 4: Synthèse des variables environnementales importantes pour la distribution du thon rouge(Benyahia et al., 2021).....	31
Tableau 5: Récapitulatif des bibliothèques python utilisées pour le traitement et la modélisation.	41
Tableau 6: Valeurs manquantes par variable environnementale.....	45
Tableau 7:Principales métriques d'évaluation des modèles de régression.	50
Tableau 8: Performances des modèles avant et après optimisation des hyperparamètres pour la prédiction de la position géographique du thon rouge (target 1).....	55
Tableau 9: Performances des modèles avant et après optimisation des hyperparamètres pour la prédiction de la longueur du thon rouge (target 2 — LenCM).....	60
Avant fine-tuning (Tableau 10 :le modèle Random Forest donne les meilleurs paramètres avec un coefficient de détermination R² de 0,8023 ainsi que les plus faibles erreurs (RMSE = 27,29 cm et MAE = 17,05 cm), indiquant une forte capacité d'estimer la taille des individus à partir des variables environnementales.....	Error! Bookmark not defined.
Tableau 11: Tableau récapitulatif du classement combiné des variables pour target 1 (Position) — Modèle XGBoost + Grid Search — Méditerranée	65
Tableau 12: Tableau récapitulatif du classement combiné des variables pour target 2 (Longueur en cm) — Modèle Random Forest Bayesian — Méditerranée.....	67

Liste des Abréviations

AI / IA	Artificiel Intelligence / Intelligence Artificielle
ANN	Artificiel Neural Networks (Réseau de Neurones Artificiels)
ASHA	Asynchrones Successive Halving Algorithm
BFT	Bluefin Tuna (Thon rouge de l'Atlantique)
CGPM / GFCM	Commission Générale des Pêches pour la Méditerranée
CHL / Chl-a	Chlorophylle-a
CMEMS	Copernicus Marine Environment Monitoring Service
CPUE	Catch Per Unit Effort (Capture Par Unité d'Effort)
DL	Deep Learning (Apprentissage profond)
EMODnet	European Marine Observation and Data Network
EPSG	European Petroleum Survey Group
GAM:	Generalized Additive Models (Modèles Additifs Généralisés)
GEBCO	General Bathymetric Chart of the Oceans
GLM	Generalized Linear Models (Modèles Linéaires Généralisés)
GPS	Global Positioning System
GBYP	Global Bluefin Tuna Year Program (programme ICCAT)
HAL	Hyper Articles en Ligne
HCMR	Hellenic Centre for Marine Research
ICM-CSIC	institut de Ciencias del Mar (Espagne)
ICCAT	International Commission for the Conservation of Atlantic Tunas
IEO-CSIC	Institut Español de Oceanografía
IFREMER	Institut Français de Recherche pour l'Exploitation de la Mer
INRH	Institut National de Recherche Halieutique (Maroc)
INSTM	Institut National des Sciences et Technologies de la Mer (Tunisie)
IRD	Institut de Recherche pour le Développement
L4	Level 4 (niveau de traitement des données satellitaires)
LSTM	Long Short-Term Memory (type de réseau de neurones)
LSSS	Larval Survey and Sampling System
MAE	Mean Absolute Error (Erreur Absolue Moyenne)

Liste des Abréviations

ML	Machine Learning (Apprentissage automatique)
MODIS	Moderate Resolution Imaging Spectroradiometer
MSE	Mean Squared Error (Erreur Quadratique Moyenne)
NEMO	Nucleus for European Modelling of the Ocean
NetCDF	Network Common Data Form (format de données)
NOAA	National Oceanic and Atmospheric Administration
NUIST	Nanjing University of Information Science and Technology
OGS	Institute Nazionale di Oceanografia e di Geofisica Sperimentale (Italie)
OLCI	Ocean and Land Colour Instrument (Sentinel-3)
PAT / PSAT	Pop-up Archival Tag / Pop-up Satellite Archival Tag (balise satellite)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
MDI	Diminution Moyenne de l'Impureté

Liste des Symboles

°C	Degré Celsius (unité de température)
Mg m⁻³	Milligramme par mètre cube (concentration de chlorophylle-a)
mmol m⁻³	Millimole par mètre cube (concentration d'oxygène dissous)
M s⁻¹	Mètre par seconde (vitesse des courants marins)
M	Mètre (profondeur, hauteur des vagues)
Km	Kilomètre (distance)
Kg	Kilogramme (poids)
Cm	Centimètre (longueur)
PSU	Practical Salinity Unit — Unité Pratique de Salinité
0,5° × 0,5°	Résolution spatiale de la grille ICCAT
R²	Coefficient de détermination (qualité d'ajustement du modèle)
R² ajusté	Coefficient de détermination ajusté au nombre de variables
N	Nombre d'observations (effectif de l'échantillon)
Y_i	Valeur réelle observée pour l'observation i
Ŷ_i	Valeur prédite par le modèle pour l'observation i
$\bar{y}$	Moyenne des valeurs observées
Λ	Lambda — paramètre de régularisation L2 dans XGBoost
α	Alpha — paramètre de régularisation L1 dans XGBoost
Q1	Premier quartile (25ème percentile)
Q3	Troisième quartile (75ème percentile)
Q3 – Q1	Intervalle interquartile (IQR) — mesure de dispersion robuste
U₀	Composante zonale (Est-Ouest) de la vitesse du courant marin
V₀	Composante méridionale (Nord-Sud) de la vitesse du courant

Résumé :

Le thon rouge (*Thunnus thynnus*) est une espèce pélagique d'une grande importance écologique et économique dans le bassin méditerranéen. Sa distribution est influencée par plusieurs facteurs environnementaux tels que la température de surface de la mer, la salinité, la chlorophylle-a, les courants marins, la bathymétrie et la hauteur de la surface de la mer. L'objectif de ce travail est de modéliser et de prédire la distribution spatio-temporelle du thon rouge à l'aide de techniques d'intelligence artificielle et d'apprentissage automatique. À partir de données halieutiques et environnementales géoréférencées, plusieurs modèles prédictifs, notamment la Régression Linéaire Multiple, Random Forest et XGBoost, ont été développés, évalués et comparés. Les résultats obtenus montrent que les modèles basés sur les arbres de décision, en particulier Random Forest et XGBoost, offrent de meilleures performances prédictives que les approches linéaires classiques pour représenter les relations complexes entre les variables environnementales et la présence du thon rouge. Cette étude a également permis d'identifier les facteurs environnementaux les plus influents ainsi que les zones potentielles de forte concentration de l'espèce dans le bassin méditerranéen. Les résultats obtenus confirment l'intérêt de l'intelligence artificielle pour améliorer la compréhension de la dynamique écologique du thon rouge et contribuer à une gestion rationnelle durable des ressources halieutiques.

Mot clé : Thon rouge, Intelligence artificielle, , Distribution spatio-temporelle, Facteurs environnementaux, , Bassin méditerranéen, Modélisation.

Abstract:

Atlantic bluefin tuna (*Thunnus thynnus*) is a pelagic species of major ecological and economic importance in the Mediterranean Basin. Its distribution is influenced by several environmental factors, including sea surface temperature, salinity, chlorophyll-a concentration, marine currents, bathymetry, and sea surface height. The objective of this study is to model and predict the spatio-temporal distribution of Atlantic bluefin tuna using artificial intelligence and machine learning techniques. Based on georeferenced fisheries and environmental datasets, several predictive models, including Multiple Linear Regression, Random Forest, and XGBoost, were developed, evaluated, and compared. The results show that tree-based models, particularly Random Forest and XGBoost, provide better predictive performance than traditional linear approaches in capturing the complex relationships between environmental variables and the presence of Atlantic bluefin tuna. This study also identified the most influential environmental factors and the potential areas of high species concentration within the Mediterranean Basin. The findings highlight the value of artificial intelligence in improving the understanding of the ecological dynamics of Atlantic bluefin tuna and supporting the sustainable management of fisheries resources.

Keywords: Atlantic bluefin tuna, Artificial Intelligence, , Spatio-temporal Distribution, Environmental Factors, Mediterranean Basin, Modeling.

ملخص :

يُعدّ التونة الحمراء الأطلسية من الأسماك السطحية الكبيرة ذات الأهمية البيئية والاقتصادية البالغة في حوض البحر ويتأثر توزيعها بعدة عوامل بيئية، من بينها درجة حرارة سطح البحر، والملوحة، وتركيز الكلوروفيل، الأبيض المتوسط. يهدف هذا العمل إلى نمذجة والتنبؤ بالتوزيع المكاني والزمني. والتيارات البحرية، وعمق المياه، وارتفاع سطح البحر واعتمادًا على بيانات المصايد والبيانات البيئية. للتونة الحمراء الأطلسية باستخدام تقنيات الذكاء الاصطناعي والتعلم الآلي. المكانية المرجعية، تم تطوير وتقييم ومقارنة عدة نماذج تنبؤية.

أظهرت النتائج أن النماذج المعتمدة على أشجار القرار حققت أداءً تنبؤيًا أفضل مقارنة بالنماذج الخطية التقليدية، وذلك كما أتاحت هذه الدراسة تحديد أهم. لقدرتها على تمثيل العلاقات المعقدة بين المتغيرات البيئية ووجود التونة الحمراء. العوامل البيئية المؤثرة وتحديد المناطق المحتملة ذات التركيز المرتفع لهذه الأنواع داخل حوض البحر الأبيض المتوسط. وتؤكد النتائج المتحصل عليها أهمية الذكاء الاصطناعي في تحسين فهم الديناميكية البيئية للتونة الحمراء الأطلسية. والمساهمة في الإدارة المستدامة للموارد السمكية.

الكلمات المفتاحية

التونة الحمراء الأطلسية، الذكاء الاصطناعي، التوزيع المكاني الزمني، العوامل البيئية، حوض البحر الأبيض المتوسط، النمذجة.

Introduction

Introduction

Les grands poissons pélagiques, notamment le thon rouge (*Thunnus thynnus*), occupent une place spéciale dans les écosystèmes marins méditerranéens. Situé au sommet de la chaîne trophique, il joue un rôle fondamental dans le maintien des équilibres écologiques des océans (Forrestal et al., 2012). Il régule les populations de ses proies et contribue au fonctionnement global des écosystèmes marins. Au-delà de son importance écologique, cette espèce représente une ressource halieutique stratégique de grande valeur économique dans plusieurs pays côtiers de la Méditerranée. Elle soutient les activités de pêche industrielle et artisanale, générant des revenus considérables et participant à la sécurité alimentaire de milliers d'habitants des villes côtières (Logan et al., 2025). Cependant, les populations de cette espèce subissent aujourd'hui de multiples pressions anthropiques, notamment la dégradation des habitats marins et les effets croissants du changement climatique (Andrews et al., 2023). Ces facteurs influencent directement la distribution spatio-temporelle de l'espèce. Ils peuvent également influencer son comportement migratoire ainsi que ses milieux de reproduction et d'alimentation (J. Fromentin et al., 2014).

Dans ce contexte, la compréhension des relations entre le thon rouge et son environnement est devenue un enjeu majeur pour la gestion durable des ressources marines et la conservation de la biodiversité. Les avancées technologiques en observation de la Terre, les systèmes d'information géographique et les bases de données halieutiques ont permis la production de volumes importants de données environnementales et biologiques. Ces données comprennent notamment la température de surface de la mer, la salinité de surface de la mer, la chlorophylle-a, la vitesse des courants marins, la hauteur des vagues, la profondeur et la distance à la côte, ainsi que des données de capture telles que la taille et le poids de l'espèce (Druon et al., 2016).

Toutefois, la complexité des interactions entre ces variables et la dynamique des populations de thon rouge rend difficile l'analyse à l'aide des approches statistiques classiques. Dans ce cadre, les méthodes d'intelligence artificielle (IA) et d'apprentissage automatique (Machine Learning) apparaissent comme des outils particulièrement prometteurs grâce à leur capacité à traiter un grand volume de données hétérogènes et à identifier des relations non linéaires, complexes et

Introduction

multidimensionnelles (Yang et al., 2025). Ces approches permettent d'améliorer la modélisation et la prédiction de la distribution spatio-temporelle de cette espèce. Elles offrent également de nouvelles perspectives pour la compréhension des mécanismes écologiques qui gouvernent la migration et la répartition de cette espèce dans le bassin méditerranéen.

La problématique centrale de ce projet de fin d'études est la suivante :

Comment les techniques d'intelligence artificielle peuvent-elles être utilisées pour modéliser et prédire la distribution spatio-temporelle des grands pélagiques dans le bassin méditerranéen en fonction des facteurs environnementaux, afin d'améliorer la compréhension de leur dynamique écologique et de contribuer à une gestion rationnelle et durable de cette ressource halieutique ?

- Analyser les relations entre les observations des grands pélagiques et les variables environnementales à partir de données géoréférencées
- Développer des modèles d'intelligence artificielle capables de modéliser la distribution spatio-temporelle des grands pélagiques
- Identifier les zones potentielles de forte concentration de cette espèce dans le bassin méditerranéen
- Valider et comparer les performances des modèles prédictifs développés
- Visualiser et interpréter les résultats afin de soutenir la gestion durable des ressources halieutiques.

L'hypothèse principale de ce travail est que l'intégration conjointe des données halieutiques et environnementales dans des modèles d'intelligence artificielle permet d'améliorer significativement la précision de la prédiction de la distribution spatio-temporelle de cette espèce dans le bassin méditerranéen.

Les résultats attendus contribueront à une meilleure compréhension des habitats favorables à cette espèce, fournissant ainsi des outils d'aide à la gestion rationnelle et durable de cette ressource marine.

Enfin, ce mémoire est structuré en plusieurs chapitres :

- Le premier chapitre : Synthèse bibliographique et généralités sur les grands pélagiques

Introduction

- Le deuxième chapitre : description des données, de la zone d'étude et de la méthodologie mise en œuvre.
- Le troisième chapitre : Résultats obtenus, leur analyse et discussions des résultats, suivis de la conclusion de l'étude et des perspectives futures.

Chapitre I : Synthèse Bibliographique et Généralité

1. Recherche bibliographique

1.1 Stratégie de recherche documentaire

La présente revue de littérature a été conduite selon une démarche systématique et transparente, visant à garantir la reproductibilité et la rigueur scientifique de la sélection bibliographique. L'ensemble de la démarche est conforme aux recommandations PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) adaptées aux sciences halieutiques et à l'écologie marine (Moher et al., 2009).

Les sources de données ont été constituées en identifiant les articles les plus pertinents extraits de cinq bases de données scientifiques, à savoir : Web of Science (WoS), Scopus, Google Scholar, HAL et ResearchGate. Tous les articles ont été organisés manuellement à l'aide de l'outil *Mendeley* et dans un fichier Excel (tableau 1).

1.1.1 Critères d'inclusion et d'exclusion

La recherche documentaire a été réalisée à partir de l'équation booléenne suivante, appliquée aux bases de données scientifiques :

("intelligence artificielle" OR "machine learning" OR "deep learning" OR "artificial intelligence") AND ("modélisation spatio-temporelle" OR "species distribution model" OR "SDM" OR "habitat model") AND ("grands pélagiques" OR "thon rouge" OR "bluefin tuna" OR "Thunnus thynnus" OR "espadon" OR "pelagic fish") AND ("Méditerranée" OR "Mediterranean basin" OR "Mediterranean Sea") AND ("facteurs environnementaux" OR "environmental variables" OR "SST" OR "chlorophyll").

Tableau 1: Étapes des critères de sélection des données.

Filtre	Critère d'inclusion	Critère d'exclusion	N restant
Sélection initiale (mots-clés)	L'équation de recherche	/	1 710
Période temporelle	2018 – 2026	Avant 2018	1 350
Type de document	Articles de recherche originaux	Revue de littérature, chapitre , thèses	950
Langue	Anglais et français	Autres langues	900

Domaine d'étude	Biologie marine, océanographie, IA appliquée, télédétection	/	500
Doublons	Tous les articles uniques	~20 doublons entre bases de données	480
Pertinence manuelle (score $\geq 8/10$)	L'IA, grands pélagiques méditerranéens, facteurs environnementaux	Articles peu pertinents (score $< 8/10$)	27

Nous avons développé une base de données relationnelle regroupant les articles étroitement liés à notre domaine de recherche. Cette base de données comprenait plusieurs tables : « Article », « Mots-clés », « Auteurs », « Source des données », « Zone d'étude », « Indicateurs », « Modèle » et « Modèle comparé ». La table « Article » constituait la table principale, dont les clés primaires étaient reliées à l'ensemble des autres tables (tableaux 1 et 2).

Tableau 2: Répartition du domaine d'étude des 27 articles sélectionnés.

Domaine thématique	Nombre d'Articles	Proportion (%)
Biologie marine et halieutique (thon rouge, pélagiques)	8	30 %
Modèles de distribution des espèces (SDM, MaxEnt, GAM)	6	22 %
Intelligence artificielle et Machine Learning (RF, XGBoost, DL)	7	26 %
Télédétection et variables environnementales	4	15 %
Gestion halieutique durable	2	7 %

Les données ont été saisies manuellement, article par article. Au total, 1710 publications ont été intégrées dans la base de données, dont 27 articles identifiés comme étant les plus pertinents pour notre thématique de recherche. Ces 27 articles provenaient des données WoS préalablement traitées et sélectionnées dans les différents domaines d'étude.

1.2. Analyse du réseau institutionnel de la recherche

Les institutions les plus influentes dans la recherche relative à notre sujet, mesurées par le nombre de citations. IFREMER (France) domine avec 245 citations, suivi d'IEO-CSIC (Espagne) et HCMR (Grèce) (tableau 3).

Tableau 3: Classement des institutions de recherche méditerranéennes par citations et thématique.

Citations	Institution	Pays	Cluster thématique
245	IFREMER	France	Halieutique & Modélisation
198	IEO-CSIC	Espagne	Océanographie opérationnelle
176	HCMR	Grèce	Écologie méditerranéenne
165	IRD	France	Halieutique & Modélisation
142	University of Genoa	Italie	Écologie méditerranéenne
128	ICM-CSIC	Espagne	Océanographie opérationnelle
115	OGS	Italie	Géophysique marine
98	INSTM	Tunisie	Recherche Sud Méditerranée
87	INRH	Maroc	Recherche Sud Méditerranée

Reflétant leur forte visibilité scientifique. Les résultats montrent également une spécialisation thématique des institutions autour de l'halieutique, de l'océanographie opérationnelle, de l'écologie méditerranéenne et de la géophysique marine. Enfin, la présence de l'INSTM (Tunisie) et de l'INRH (Maroc) souligne la contribution croissante des institutions du sud de la Méditerranée à la recherche régionale (tableau 3).

1.3. Synthèse basée sur les modèles d'IA

L'analyse basée sur les modèles d'IA montre que les approches d'apprentissage automatique (Random Forest, XGBoost et CNN) sont largement dominantes dans les études menées dans les océans Atlantique, Pacifique et Indien, alors que leur utilisation reste plus limitée en Méditerranée. À l'inverse, les modèles écologiques tels que GAM et MaxEnt sont davantage employés dans le bassin méditerranéen pour la modélisation

spatio-temporelle de la distribution et de l'habitat des grands poissons pélagiques (figure 1).

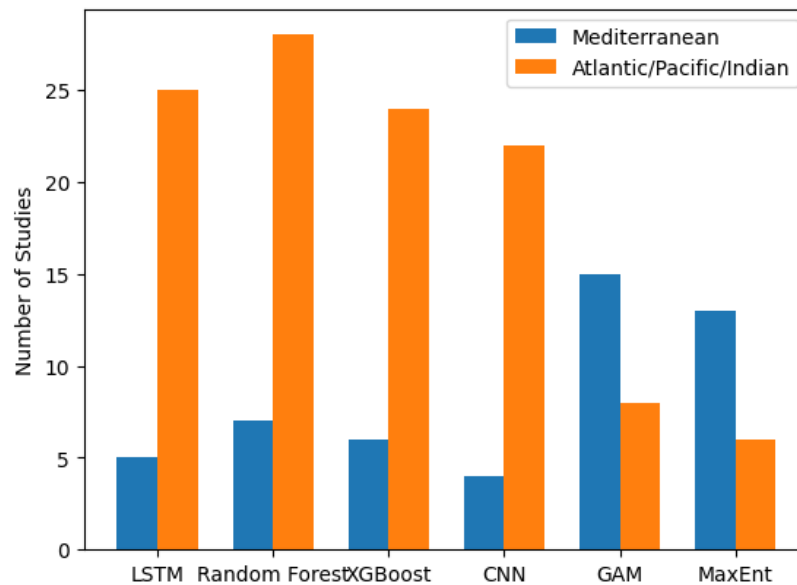


Figure 1: Nombre d'études utilisant différents modèles d'IA dans la région méditerranéenne et dans les océans Atlantique, Pacifique et Indien.

Cette différence reflète une transition progressive vers les méthodes d'intelligence artificielle dans les études intégrant les variables environnementales.

1.4. Synthèse des résultats bibliométriques et méthodologiques

L'analyse du corpus de 27 références met en évidence plusieurs tendances majeures concernant l'application de l'intelligence artificielle à la modélisation de la distribution des grands poissons pélagiques. La majorité des études se concentre sur l'évaluation des relations entre la répartition des espèces et les conditions océanographiques, notamment la température de surface de la mer (SST), la concentration en chlorophylle-a (Chl-a) et la dynamique des courants marins. Les résultats montrent que les approches d'apprentissage automatique, telles que Random Forest, Gradient Boosting et les réseaux de neurones, offrent généralement de meilleures performances prédictives que les méthodes statistiques classiques comme les modèles linéaires généralisés (GLM) et les modèles additifs généralisés (GAM). Toutefois, l'utilisation spécifique de ces techniques d'intelligence artificielle pour la modélisation spatio-temporelle des thons rouges et de la distribution des espèces en Méditerranée demeure encore limitée. Par ailleurs, peu d'études ont réalisé des comparaisons systématiques et robustes entre différents modèles d'IA à l'échelle de l'ensemble du bassin méditerranéen, ce qui restreint l'identification des approches les plus adaptées. Enfin, l'intégration conjointe

de multiples sources de données, telles que la télédétection satellitaire, les données de pêche et les informations issues du marquage des espèces, reste insuffisamment développée, malgré son potentiel pour améliorer la précision et la robustesse des modèles prédictifs.

2. Le grand pélagique (*Thunnus thynnus*)

2.1. Importance écologique, économique et pression anthropique

a) Écologique

Parmi les espèces emblématiques de la Méditerranée figurent les grands poissons pélagiques, le thon rouge de l'Atlantique (*Thunnus thynnus*), (Reglero et al., 2012). Cette espèce hautement migratrice occupe une position clé au sommet de la chaîne trophique et contribue à l'équilibre des écosystèmes marins (Bănar et al., 2010).

b) Economique

En raison de sa forte valeur économique et de la demande internationale, notamment sur le marché du sushi et du sashimi, le thon rouge a fait l'objet d'une exploitation intense au cours des dernières décennies. Les grands pélagiques représentent une ressource halieutique stratégique à l'échelle mondiale (J. M. Fromentin & Powers, 2005).

c) Pression anthropique

Les populations du thon rouge sont aujourd'hui confrontées à plusieurs pressions anthropiques, notamment la surexploitation, la dégradation des habitats marins et les effets du changement climatique, qui peuvent modifier leur dynamique et leur distribution spatiale (Karp et al., 2025). Cette pression de pêche intense a conduit à la mise en place de mesures de gestion et de conservation par la Commission Internationale pour la Conservation des Thonidés de l'Atlantique (ICCAT) afin de garantir la durabilité des stocks (Sumaila & Huang, 2012).

3. Biologie et écologie du thon rouge (*Thunnus thynnus*)

3.1. Taxonomie et adaptations physiologiques

a) Taxonomie

Selon (Collette et al., 2001; J. M. Fromentin & Powers, 2005), le thon rouge est classé comme suit :

- **Embranchement** : Chordés
- **Sous-embranchement** : Vertébrés
- **Superclasse** : Gnathostomes
- **Classe** : Ostéichthyens
- **Sous-classe** : Actinoptérygiens
- **Ordre** : Perciformes
- **Sous-ordre** : Scombroïdes
- **Famille** : Scombridés
- **Genre** : Thunnus
- **Espèce** : Thunnus thynnus (Linnaeus, 1758).

b) Adaptations physiologiques

Sur le plan physiologique, le thon rouge possède une capacité remarquable d'endothermie régionale, lui permettant de maintenir certaines parties de son corps à une température supérieure à celle de l'eau environnante (J. M. Fromentin & Powers, 2005). Cette adaptation améliore ses performances locomotrices et lui permet d'exploiter une grande diversité d'habitats marins, allant des eaux froides tempérées aux eaux plus chaudes des zones de reproduction. Grâce à ces adaptations, l'espèce peut tolérer une large gamme de températures, approximativement entre 3 °C et 30 °C (J. M. Fromentin & Powers, 2005).

3.2 Morphologie

Les adultes peuvent mesurer de 3 à 4 mètres et peser plus de 700 kg, avec certains individus qui pourraient atteindre près de 900 kg (Benyahia et al., 2021). Son corps de forme fuselée et hydrodynamique est conçu pour une nage rapide et constante. La morphologie se distingue en particulier par la présence de deux nageoires dorsales distinctes et une forte nageoire caudale en forme de croissant (Figure 2).

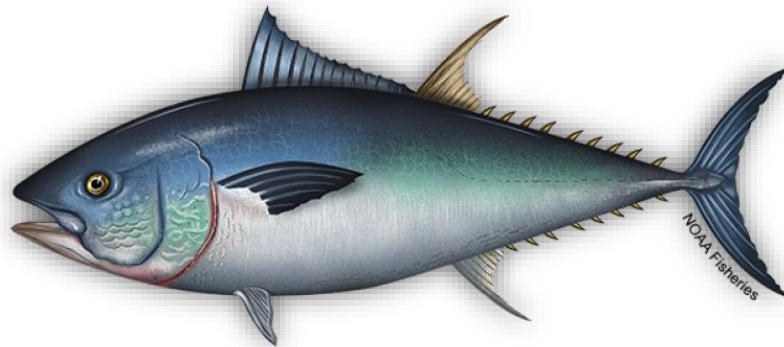


Figure 2: Thon rouge de l'Atlantique (*Thunnus thynnus*) (NOAA, 2026).

3.3 Reproduction migrations et utilisation de l'habitat en Méditerranée

a) Reproduction

La reproduction du thon rouge est liée à des conditions environnementales spécifiques, telles que des températures de surface assez hautes et des configurations océanographiques particulières. Par exemple, dans le golfe du Mexique, les thons rouges adultes choisissent des habitats présentant des variations de température, la présence de courants et des zones d'activité océanique intense, des éléments qui semblent contribuer à la réussite de leur reproduction (Teo et al., 2007).

b) migrations et utilisation de l'habitat en Méditerranée

Le thon rouge est une espèce pélagique largement distribuée dans l'océan Atlantique Nord ainsi que dans ses mers adjacentes, notamment la mer Méditerranée. Sa répartition géographique s'étend globalement de l'équateur jusqu'aux eaux tempérées du nord de l'Atlantique et inclut la mer Méditerranée ainsi que certaines zones du golfe du Mexique (Benyahia et al., 2021).

Le thon rouge est connu pour ses migrations saisonnières de grande amplitude entre les zones d'alimentation situées dans les régions tempérées de l'Atlantique et les zones de reproduction localisées dans des eaux plus chaudes. Deux principales zones de reproduction sont distinguées : le golfe du Mexique pour le stock occidental et la mer

Méditerranée pour le stock oriental (Benyahia et al., 2021). Du point de vue écologique, le thon rouge occupe principalement les eaux pélagiques de surface et de subsurface. Les études basées sur le marquage électronique ont montré que les individus peuvent effectuer des plongées profondes atteignant parfois 500 à 1000 mètres, généralement associées à la recherche de nourriture ou à des mécanismes de régulation thermique (Druon et al., 2016) (figure 3).

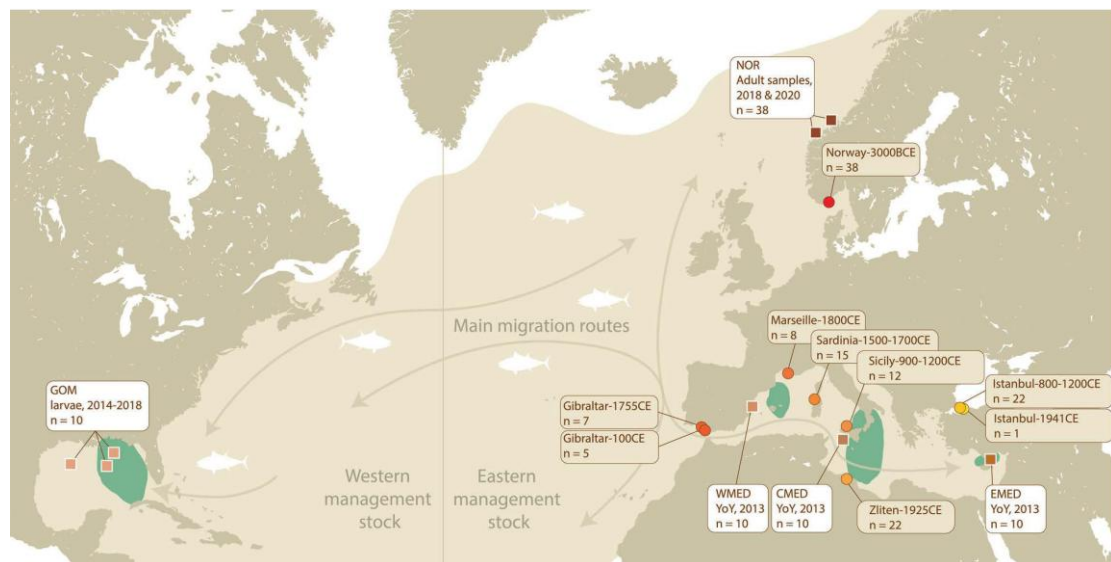


Figure 3: Répartition spatiale, zones de reproduction (en vert) et principales voies migratoires du thon rouge (*Thunnus thynnus*) entre la Méditerranée et l'Atlantique Est (Eriksen et al., 2025)

4. Facteurs environnementaux contrôlant la distribution

4.1 Variables physiques

La distribution des espèces pélagiques est fortement influencée par les conditions environnementales et océanographiques. Des paramètres tels que : La température de surface de la mer (SST), la chlorophylle-a ou les courants marins déterminent en grande partie les zones d'alimentation, de reproduction et de migration de ces espèces. Plusieurs études ont montré que la dispersion des larves de thon dépend étroitement de ces facteurs environnementaux, notamment de la température de l'eau et de la productivité primaire (Russo et al., 2021).

Les facteurs océanographiques sont essentiels dans le choix des habitats propices au thon rouge de l'Atlantique. La distribution spatiale et temporelle peut être affectée par des facteurs tels que la température de surface marine, les gradients de chlorophylle et les irrégularités dans la hauteur de la mer (Druon et al., 2016).

4.2 Variables biogéochimiques

- La chlorophylle-a, qui représente le premier élément de la chaîne alimentaire en milieu marin, sert d'indicateur indirect pour localiser les zones où se concentrent les thons rouges (Benyahia et al., 2021). Dans les zones de surabondance, les concentrations moyennes demeurent faibles, oscillant entre 0,11 et 0,14 (mg m^{-3}) au printemps et entre 0,15 et 0,23 (mg m^{-3}) durant l'été (Teo et al., 2007).
- La salinité est une caractéristique hydrologique essentielle qui décrit les masses d'eau. Les régions où le thon rouge est particulièrement productif affichent une salinité stable, oscillant entre 36,6 et 37,1 (PSU) au printemps et entre 37,8 et 37,9 (PSU) en été (tableau 4).
- L'oxygène dissous est crucial pour le métabolisme aérobie actif du thon rouge. Les concentrations dans les zones de prise fluctuent entre 235,6 et 243 (mmol m^{-3}) au printemps, et entre 229 et 235,4 (mmol m^{-3}) en été. Il y a une relation inverse entre l'oxygène et la température ($r = -0,64$), et on considère que l'oxygène limite la répartition des thonidés (Muhling et al., 2017).

Tableau 4: Synthèse des variables environnementales importantes pour la distribution du thon rouge (Benyahia et al., 2021).

Variable	Unité	Gamme favorable (Thunnus thynnus)	Source de données
Température de surface (SST)	°C	18 – 26 °C (reproduction) ; 10 – 28 °C (alimentation)	MODIS, Sentinel-3 (Copernicus)
Chlorophylle-a	mg m^{-3}	0,10 – 0,23 mg m^{-3}	MODIS-Aqua, OLCI Sentinel-3
Salinité	PSU	36,6 – 37,9 PSU	CMEMS (modèle NEMO)
Oxygène dissous	mmol m^{-3}	229 – 243 mmol m^{-3}	CMEMS biogéochimique
Hauteur de la mer (SSH)	m	Gradients / anomalies	CMEMS altimétrie
Bathymétrie	m	0 – 1000 m (habituellement)	GEBCO, EMODnet

4.3 Données mobilisables pour l'étude des pélagiques

La modélisation de la dynamique spatio-temporelle des grands pélagiques, notamment du thon rouge (*Thunnus thynnus*), repose en grande partie sur la capacité des méthodes d'intelligence artificielle à exploiter des données nombreuses et de nature variée (Precioso et al., 2022). Cette approche s'appuie sur la combinaison de deux types d'informations complémentaires.

4. 4 Données environnementales et océanographiques (télédétection, SST, chlorophylle)

Pour comprendre où se trouvent les grands pélagiques, il faut d'abord caractériser le milieu dans lequel ils évoluent. La télédétection satellitaire constitue aujourd'hui la source de données la plus efficace car elle permet de suivre des paramètres comme la température de surface (SST) ou la chlorophylle sur de grandes distances, avec une précision et une régularité que des relevés en mer ne pourraient pas atteindre (Benyahia et al., 2021; Druon et al., 2016). Dans les modèles de distribution, ces indicateurs sont généralement classés en deux groupes. On trouve d'abord les variables statiques, qui définissent le cadre physique de l'habitat (bathymétrie, distance à la côte). À côté de cela, les variables dynamiques permettent de suivre les changements du milieu en temps réel (température, courants, fronts thermiques) (Russo et al., 2021).

4.5. Données biologiques et halieutiques (captures, campagnes scientifiques, marquage)

L'identification des zones préférentielles d'une espèce nécessite des données de terrain robustes. Pour les grands pélagiques comme le thon rouge (*Thunnus thynnus*), la source d'information historique la plus volumineuse provient des statistiques de pêche centralisées par l'ICCAT (ICCAT, 2025). Ces bases de données offrent l'avantage d'une couverture géographique complète à l'échelle du bassin et de longues séries temporelles permettant de capter des variations de distribution sur plusieurs décennies (J. M. Fromentin & Powers, 2005).

5. Modélisation de la distribution, intelligence artificielle et données

5.1 Modèles de distribution des espèces : principes et limites des approches classiques

L'étude de la répartition des espèces repose sur le concept de niche écologique développé selon la théorie de Hutchinson. Chaque espèce vit dans un environnement

défini par plusieurs facteurs environnementaux, comme la température, la salinité ou la nourriture disponible. Dans les modèles de distribution des espèces (MDE), on distingue la niche fondamentale, qui représente les conditions idéales de vie d'une espèce, et la niche réalisée, qui correspond aux conditions réellement occupées dans la nature (Elith & Leathwick, 2009).

Pendant de nombreuses années, la modélisation des ressources marines s'est principalement appuyée sur les modèles linéaires généralisés (MLG) et les modèles additifs généralisés (AGM). Ces méthodes ont longtemps été considérées comme des outils de référence. Cependant, elles présentent aujourd'hui certaines limites, notamment pour représenter les relations complexes entre les variables environnementales et pour traiter les grandes quantités de données satellitaires disponibles actuellement (Karp et al., 2025).

5.2 Apprentissage automatique

L'intégration de l'apprentissage automatique (Machine Learning) dans l'étude des grands pélagiques représente une véritable rupture conceptuelle. Alors que les approches traditionnelles s'efforcent principalement d'expliquer des relations complexes, le Machine Learning se focalise sur la maximisation de la précision prédictive (Kühn et al., 2025).

5.3 Données mobilisables : télédétection, pêcheries et intégration multi-sources

La mise en œuvre opérationnelle des algorithmes d'intelligence artificielle repose sur une des données où la télédétection et les relevés halieutiques ne sont plus de simples sources isolées, mais des flux fusionnés au sein d'un même système cohérent. Les données satellites agissent comme une grille de prédicteurs spatialisés, tandis que les données de pêche fournissent la « vérité terrain » indispensable à l'apprentissage supervisé (Druon et al., 2016).

6. Positionnement du projet

6.1. Lacunes scientifiques identifiées

Malgré la richesse des travaux consacrés aux grands pélagiques en Méditerranée occidentale, l'analyse de la littérature révèle plusieurs verrous scientifiques :

- Cloisonnement des sources de données : la majorité des études se concentre sur des variables physiques isolées ou des données de pêche historiques, sans intégration d'une dynamique multi-sources
- Utilisation encore limitée des méthodes numériques avancées en Méditerranée
- Manque de comparaisons systématiques entre différents algorithmes d'IA au sein d'une même étude
- Absence d'outils de gestion opérationnels, dynamiques et directement exploitables par les décideurs

6.2. Apports attendus de l'étude

Pour combler les lacunes définies précédemment, cette recherche propose un cadre méthodologique innovant, axé sur l'intelligence artificielle appliquée à la Méditerranée occidentale. L'objectif premier est de gérer les relations complexes entre les observations de terrain et les variables environnementales à partir de données géoréférencées de haute précision.

Chapitre II : Matériels et Méthodes

1. Zone d'étude

1.1. Présentation générale du bassin méditerranéen

La mer Méditerranée est une mer semi-fermée située entre l'Europe, l'Afrique et l'Asie. Elle communique avec l'océan Atlantique par le détroit de Gibraltar et avec la mer Noire par les détroits des Dardanelles et du Bosphore. Elle s'étend approximativement entre les longitudes 6°W et 36°E, et les latitudes 30°N et 46°N. La Méditerranée est divisée en deux grands bassins : le bassin occidental et le bassin oriental, séparés par le détroit de Sicile. Elle comprend plusieurs mers régionales importantes, telles que la mer d'Alboran, la mer des Baléares, le Golfe du Lion, la mer Tyrrhénienne, la mer Adriatique, la mer Ionienne et la mer Égée. Cette région marine possède une grande diversité écologique et joue un rôle essentiel dans les échanges climatiques, biologiques et économiques. Elle constitue également un habitat stratégique pour de nombreuses espèces marines, notamment les grands pélagiques comme les thons (Millot & Taupier-Letage, 2005).

Sur le plan hydrologique, ce sous-bassin est directement alimenté par les eaux atlantiques de surface qui pénètrent en Méditerranée via le détroit de Gibraltar — eaux relativement froides et peu salées, avant de se réchauffer et de se concentrer en sel au contact de l'atmosphère méditerranéenne. La salinité de surface y est généralement comprise entre 36 et 38 (PSU), légèrement inférieure à celle du bassin oriental. L'évaporation y dépasse largement les précipitations et les apports fluviaux, engendrant un bilan évaporatoire négatif caractéristique. Sur le plan climatique, la Méditerranée présente une variabilité saisonnière marquée de la température de surface (SST), oscillant entre 13 °C en hiver et dépassant 26 °C en été dans certaines zones côtières. Le Golfe du Lion se distingue par une forte variabilité thermique liée aux vents régionaux (mistral, tramontane), favorisant des remontées d'eaux profondes riches en nutriments. Ces conditions font de ce sous-bassin une zone de productivité primaire élevée, propice à la concentration du thon rouge (*Thunnus thynnus*) (figure 4).

Du point de vue halieutique, la Méditerranée joue un rôle stratégique majeur pour le stock Est-Atlantique et méditerranéen de thon rouge. Le Golfe du Lion, les eaux espagnoles, les eaux du Maroc et la mer Tyrrhénienne constituent des zones importantes de reproduction et d'alimentation, particulièrement actives entre mai et

août lors des migrations reproductrices des adultes depuis l'Atlantique Nord (Milot & Taupier-Letage, 2005).

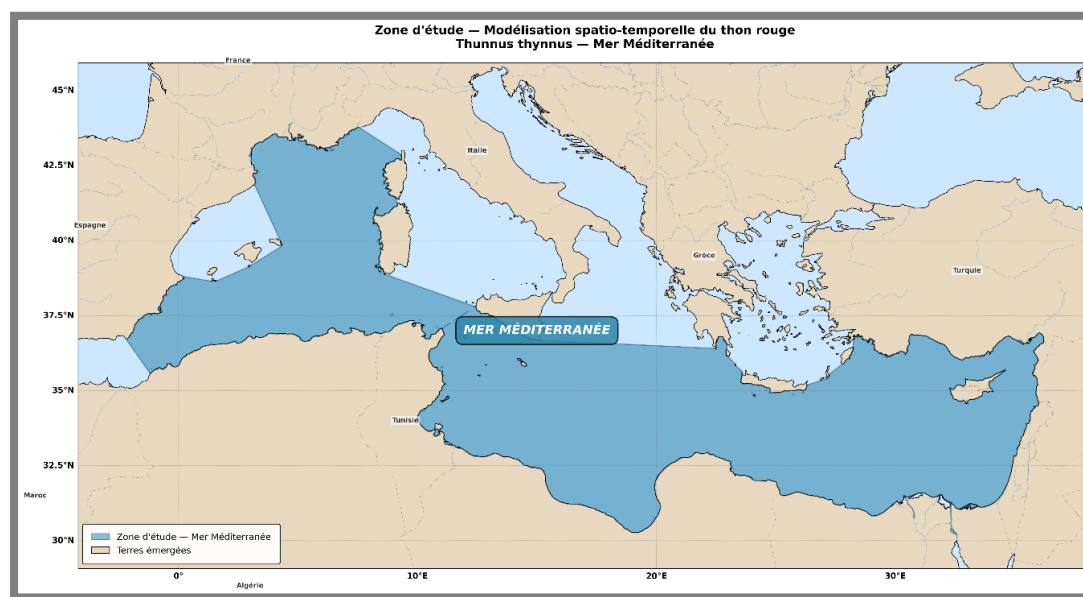


Figure 4: Zone d'étude (bassin méditerranéen).

1.1.1 Propriétés climatiques et hydrologiques

a) Propriétés climatiques

Le bassin méditerranéen se distingue par un climat de type méditerranéen classique, caractérisé par des étés chauds et secs et des hivers doux et humides. Ces conditions climatiques se reflètent directement sur la température de surface de la mer (SST), qui présente une variabilité saisonnière marquée, oscillant entre 13 °C en hiver et dépassant 28 (°C) en été dans certaines zones côtières. Au-delà de cette variabilité naturelle, la Méditerranée est aujourd'hui l'une des régions marines les plus touchées par le réchauffement climatique. Selon le Service de Surveillance Marine Copernicus (CMEMS), l'année 2024 a enregistré la température de surface annuelle la plus élevée jamais mesurée dans ce bassin, avec une moyenne de 21,5(°C), soit un excès de 1,2 (°C) par rapport à la période de référence 1991-2020 et 0,3 (°C) au-dessus du précédent record établi en 2023. Cette tendance au réchauffement, estimée entre 0,035 et 0,041 (°C) par an depuis plusieurs décennies, exerce une pression croissante sur les habitats des espèces hautement migratoires telles que le thon rouge (*Thunnus thynnus*), en modifiant progressivement les zones thermiques favorables à leur alimentation (Pisano et al., 2020).

a) Propriétés hydrologiques

La Méditerranée présente un bilan évaporatoire négatif, l'évaporation y dépasse largement les précipitations et les apports fluviaux, ce qui en fait une mer à salinité élevée, généralement comprise entre 36 et 39 (PSU) selon les sous-bassins. Cette concentration en sel influe directement sur la densité des masses d'eau et conditionne la circulation thermo-haline générale du bassin. La dynamique des eaux méditerranéennes est également marquée par des échanges permanents avec l'océan Atlantique via le détroit de Gibraltar, où des eaux atlantiques de surface plus froides et moins salées pénètrent en Méditerranée, tandis que des eaux méditerranéennes plus denses et plus salées s'écoulent en profondeur vers l'Atlantique. Ces mouvements de masses d'eau jouent un rôle structurant dans la distribution des ressources trophiques et donc dans la localisation des agrégations de grands pélagiques à travers le bassin (Millot & Taupier-Letage, 2005).

1.2 Base de données

1.2.1 Données biologique

L'analyse se concentre principalement sur les données d'observation du thon rouge (*Thunnus thynnus*) géoréférencées, couvrant la période allant de 2000 à 2019, avec un total d'environ 32 000 lignes et une résolution spatiale de $0,5^\circ \times 0,5^\circ$ à l'échelle du bassin méditerranéen occidental. Ces données sont issues de la Commission Internationale pour la Conservation des Thonidés de l'Atlantique (ICCAT), et sont associées à un ensemble de variables environnementales extraites de plateformes satellitaires et de modèles océanographiques tels que le service Copernicus (CMEMS). Cette résolution spatiale est cohérente avec les standards adoptés dans les publications récentes portant sur la modélisation spatio-temporelle de la distribution des grands pélagiques en Méditerranée ([Díaz-Barroso et al., 2022](#)).

1.2.2 Données environnementales

Les données environnementales utilisées pour l'entraînement des modèles couvrent la période de 2000 à 2019 et proviennent principalement du service Copernicus Marine Environment Monitoring Service (CMEMS). Ces données sont extraites à partir de produits satellitaires et de réanalyse océanique, avec une résolution spatiale de $0,5^\circ \times$

0,5° cohérente avec la grille ICCAT des observations de pêche. L'ensemble des variables environnementales comprend : la température de surface de la mer (SST) en degrés Celsius, la salinité de surface (SSS) en PSU, la concentration en chlorophylle-a (CHL) en $\text{mg}\cdot\text{m}^{-3}$, la vitesse des courants marins (current_speed) en m/s, la hauteur significative des vagues (wave_height) en mètre, ainsi que la bathymétrie (depth) en mètres et la distance à la côte (distance_to_coast) en kilomètres. Ces variables ont été sélectionnées sur la base de leur pertinence écologique documentée dans la littérature pour la modélisation de la distribution du thon rouge (*Thunnus thynnus*) en Méditerranée. Au total, 32 537 observations validées ont été extraites après contrôle de qualité, filtrage spatial et complétion des valeurs manquantes via les produits satellitaires gap-free de Copernicus. pour la phase de validation , les données environnementales utilisées proviennent du service Copernicus Marine (CMEMS) et couvrent l'intégralité du bassin méditerranéen (occidental et oriental) sur une période de six ans, allant du 1er janvier 2020 au 31 décembre 2025.), avec un total d'environ 6 million lignes chaque ligne représentant une combinaison unique de coordonnées géographiques (grille haute résolution) et de date journalière, avec les variables environnementales suivantes : température de surface (SST), salinité (SSS), chlorophylle-a (CHL), vitesse des courants, hauteur des vagues, distance à la côte et bathymétrie ([Azorin-Molina et al., 2023](#)).

1.2.3 Sources des données complémentaires

En complément des sources satellitaires, des données d'observations in situ issues d'océanographie halieutique ont été utilisées (Figure 5):

- Données de marquage électronique (pop-up satellite archival tags, PAT) : fournies par l'ICCAT-GBYP, permettent de retracer les trajectoires individuelles du thon rouge et de documenter son comportement thermique et de plongée ;
- Données d'indices larvaires LSSS (Larval Indices Coordination) : collectées lors de campagnes scientifiques en Méditerranée, constituant des indices de reproduction indépendants des données de pêche commerciale ;
- Données d'observations aériennes et acoustiques : issues des programmes de surveillance de la Commission Générale des Pêches pour la Méditerranée (CGPM/GFCM).

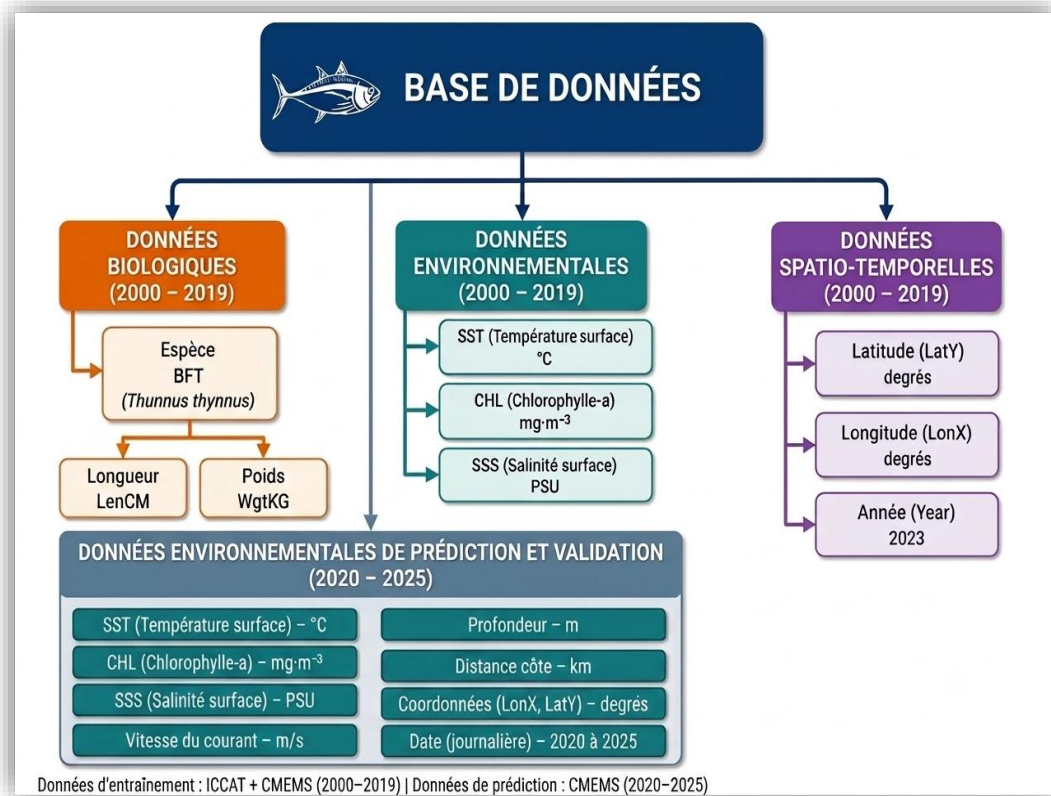


Figure 5: Schéma conceptuel des données utilisées pour la modélisation de *Thunnus thynnus* en Méditerranée.

1.3 Description des variables

1.3.1 Variables dépendantes

La variable dépendante correspond à la mesure de l'abondance relative du thon rouge (*Thunnus thynnus*), représentée sous deux modalités en fonction de la méthode de modélisation (Venolia et al., 2020).

1.3.2 Variables indépendantes (facteurs prédictifs environnementaux)

Les facteurs environnementaux susceptibles de décrire et de prévoir la répartition spatio-temporelle du thon rouge sont les variables indépendantes employées. Selon les publications récentes au sujet de la modélisation des habitats des thonidés à l'aide d'apprentissage automatique... les variables prédictives sélectionnées sont les suivantes: Température de surface, Chlorophylle-a, Salinité de surface, Hauteur de surface, Oxygène dissous, Bathymétrie, Coordonnées (Longitude/Latitude) et éléments temporels (Liu et al., 2025; Yang et al., 2025).

1.4 Outils de programmation et logiciel

1.4.1 Logiciel (Python)

Tout le processus de traitement des données, de création de modèles et de représentation graphique a été effectué en utilisant Python. On opte pour ce choix en raison de l'abondance de son écosystème de bibliothèques spécialisées dans le domaine des sciences des données pélagiques et de l'apprentissage automatique (Machine Learning). Le cadre de développement employé était Jupyter Notebook / VS Code, offrant une méthode de programmation interactive et reproductible.

1.4.2 Bibliothèque IA

L'analyse repose sur un ensemble de bibliothèques Python complémentaires et reconnues dans la communauté scientifique ([Raschka et al., 2020](#)). Le traitement et la préparation des données ont été réalisés avec Pandas et NumPy, tandis que la lecture des données océanographiques au format NetCDF a été assurée par Xarray et NetCDF4. Les modèles d'apprentissage automatique ont été implémentés via Scikit-learn pour les algorithmes classiques (Random Forest, SVM, Régression Linéaire) et XGBoost pour l'optimisation des performances prédictives (tableau 6).

Tableau 5: Récapitulatif des bibliothèques python utilisées pour le traitement et la modélisation.

Catégorie	Bibliothèque	Rôle
Traitement des données	Pandas, NumPy	Nettoyage et préparation des données
Données océanographiques	Xarray, NetCDF4	Lecture des données SST et Chl-a
Machine Learning	Scikit-learn	Régression Linéaire, Random Forest, SVM
Machine Learning	XGBoost	Amélioration des performances de prédiction
Interprétabilité	SHAP	Importance des variables explicatives

Visualisation	Matplotlib, Seaborn	Graphiques et matrices de corrélation
Cartographie	Cartopy	Cartes géospatiales du bassin méditerranéen

L'interprétabilité des résultats a été assurée par SHAP, permettant de quantifier la contribution de chaque variable environnementale. Enfin, Matplotlib et Seaborn ont été utilisés pour la visualisation statistique, et Cartopy pour la cartographie géospatiale du bassin méditerranéen occidental ([Raschka et al., 2020](#)).

1.5 Description des algorithmes utilisés

a) La Régression Linéaire Multiple

La régression linéaire multiple constitue le point de départ naturel de toute démarche de modélisation en halieutique. Son principe est simple : il suppose qu'il existe une relation proportionnelle et additive entre les variables environnementales mesurées et la variable cible. Dans le contexte méditerranéen, des travaux antérieurs ont montré que des paramètres physico-chimiques tels que la température, la salinité et l'oxygène dissous expliquent une part significative de la variabilité biologique observée chez *Thunnus thynnus* l'espèce ([Benyahia et al., 2021](#)). Ce modèle est donc utilisé dans notre étude comme **référence de base** (*baseline*), permettant d'évaluer le gain réel apporté par les algorithmes plus complexes.

Cependant, sa principale limite réside dans l'hypothèse de linéarité qu'il impose, une hypothèse rarement vérifiée dans les écosystèmes marins où les interactions entre variables sont souvent non-linéaires et multidimensionnelles ([Elith & Leathwick, 2009](#)).

b) La Forêt Aléatoire (Random Forest)

La Forêt Aléatoire est un algorithme d'ensemble qui repose sur la construction simultanée d'un grand nombre d'arbres de décision. Chaque arbre est entraîné sur un sous-ensemble aléatoire des données et des variables, puis l'ensemble des prédictions

individuelles est agrégé, par majoritaire en classification ou par moyenne en régression, pour produire une prédiction finale stable et robuste.

En océanographie, les variables environnementales sont fréquemment corrélées entre elles c'est notamment le cas de la SST et de la salinité en Méditerranée. La Forêt Aléatoire c'est la transformation préalable des données (Kühn et al., 2025). De plus, elle fournit une mesure directe de l'importance des variables, ce qui permet d'identifier quels facteurs environnementaux influencent le plus la distribution du thon rouge, un résultat précieux pour l'interprétation écologique (Effrosynidis et al., 2020).

c) Le Gradient Boosting Extrême (XGBoost)

Contrairement à la Forêt Aléatoire qui construit ses arbres en parallèle, XGBoost adopte une logique séquentielle : chaque nouvel arbre est entraîné spécifiquement pour corriger les erreurs résiduelles de l'arbre précédent. Ce processus itératif de correction, combiné à des mécanismes de régularisation intégrés (paramètres λ et α), permet d'obtenir un modèle à la fois précis et résistant au sur-apprentissage. XGBoost s'est imposé comme l'algorithme le plus performant dans plusieurs études comparatives portant sur les espèces pélagiques méditerranéennes. L'étude d'Effrosynidis et al. (2020), qui a comparé dix algorithmes différents sur huit espèces dont *Thunnus thynnus*, a conclu à sa supériorité globale pour la production de cartes d'habitat à haute résolution. Un autre avantage déterminant pour notre étude est sa capacité native à gérer les valeurs manquantes, un problème récurrent dans les données satellitaires affectées par la couverture nuageuse (Precioso et al., 2022).

d) La Machine à Vecteurs de Support (SVM)

Le SVM cherche à identifier un hyperplan optimal qui sépare au mieux les observations de deux classes, ici, présence et absence du thon rouge, en maximisant la marge entre les points les plus proches de chaque classe. Grâce à l'utilisation de fonctions noyau (kernels), cet hyperplan peut être défini dans un espace mathématique de grande dimension, ce qui permet au modèle de capturer des relations non-linéaires complexes que les méthodes classiques ne peuvent pas détecter.

Le SVM est reconnu pour sa robustesse dans les situations où le jeu de données est de taille modérée et où les classes sont déséquilibrées, une situation fréquente dans les données de présence/absence d'espèces marines (Effrosynidis et al., 2020).

2. Méthodologie

2.1 Le prétraitement des données

Le prétraitement des données constitue une étape fondamentale qui conditionne directement la qualité et la fiabilité des modèles d'intelligence artificielle développés dans ce travail. Les données brutes issues de l'ICCAT présentaient des lacunes importantes pour plusieurs variables environnementales, principalement dues à un problème structurel bien connu en océanographie côtière : le masque terrestre (Land Mask) des modèles physiques, qui génère des valeurs manquantes lorsque les coordonnées GPS d'une observation tombent à proximité immédiate du littoral. Face à ce constat, une stratégie de complétion rigoureuse a été développée, reposant exclusivement sur des données satellitaires réelles issues du service Copernicus Marine (CMEMS), sans recours à l'imputation statistique classique.

Le jeu de données brut comprenait 45 116 observations de captures de *Thunnus thynnus* en Méditerranée, couvrant la période 2000–2019. Un diagnostic systématique a révélé des taux de valeurs manquantes significatifs pour plusieurs variables clés, atteignant jusqu'à 33 % pour certains paramètres. Le principe directeur de notre approche a été le suivant : conserver toutes les valeurs originales existantes et ne compléter que les cases réellement vides, garantissant ainsi l'intégrité des données de terrain tout en assurant la complétude du dataset final.

2.2.1 Diagnostic initial des données manquantes

Avant d'engager toute procédure de remplissage, un audit systématique du jeu de données brut a été réalisé afin d'identifier précisément l'étendue et la nature des lacunes. Cette étape préliminaire est indispensable car elle conditionne le choix de la stratégie de complétion la plus adaptée à chaque variable (Maunder & Punt, 2004).

Tableau 6: Valeurs manquantes par variable environnementale.

Variable	Unité	Valeurs présentes	Valeurs manquantes	Taux de lacunes
SST (Température de surface)	°C	31 472	13 644	30,24 %
SSS (Salinité de surface)	PSU	31 472	13 644	30,24 %
Vitesse des courants	m/s	31 484	13 632	30,21 %
Hauteur des vagues	m	30 105	15 011	33,27 %
Chlorophylle-a	mg/m ³	43 234	1 882	4,17 %
Profondeur	m	45 115	1	0,002 %

L'analyse a révélé que les valeurs manquantes ne sont pas distribuées aléatoirement dans le dataset. Elles se concentrent systématiquement sur les mêmes observations, celles dont les coordonnées GPS se situent à proximité immédiate du littoral méditerranéen. Ce phénomène est directement lié au masque terrestre (Land Mask) appliqué par les modèles océaniques de Copernicus : lorsqu'un point de capture tombe dans une cellule de grille partiellement terrestre, le modèle physique renvoie automatiquement une valeur nulle, même si la pêche a réellement eu lieu en mer. Ce problème est particulièrement fréquent avec les données ICCAT dont la résolution spatiale de $0,5^\circ \times 0,5^\circ$ peut placer le centre d'une cellule sur terre alors que la zone de pêche correspondante est marine (tableau 7).

2.2.2 Le remplissage des variables environnementales manquantes

Les variables environnementales présentaient un taux élevé de valeurs manquantes, principalement dans les zones côtières. Afin de compléter ces lacunes, plusieurs produits satellitaires et de réanalyse issus du service Copernicus Marine ont été utilisés.

Pour la température de surface (SST), le produit satellitaire L4 Gap-Free de l'IFREMER (cmems_SST_MED_SST_L4_REP_OBSERVATIONS_010_021) a été

sélectionné grâce à sa couverture complète du bassin méditerranéen et à son interpolation multi-capteurs limitant les lacunes côtières. Une stratégie d'extraction annuelle en mode *lazy* a permis de réduire fortement le coût computationnel. Les valeurs manquantes ont été récupérées par sélection du pixel satellite le plus proche puis converties de Kelvin en Celsius. Après une seconde extraction utilisant un voisinage élargi, le taux de complétion a atteint 100 %.

La salinité de surface (SSS), la vitesse des courants marins et la hauteur significative des vagues ont été extraites à partir des produits globaux de réanalyse CMEMS. Une approche de voisinage valide progressif a été appliquée afin de récupérer les valeurs absentes dans les zones côtières. La vitesse des courants a été calculée à partir des composantes zonale (u_o) et méridionale (v_o) selon la relation vectorielle classique :

$$\text{Vitesse du courant} = \sqrt{(u_o^2 + v_o^2)}$$

Pour la chlorophylle-a (CHL), le produit Gap-Free multi-capteurs CMEMS a permis une extraction vectorisée directe sans recours à la stratégie de voisinage, grâce à l'absence quasi totale de lacunes dans ce produit.

Après le remplissage, toutes les variables environnementales ont atteint un taux de remplissage de 100 %. Les valeurs obtenues sont restées cohérentes avec les gammes connues pour le bassin méditerranéen selon la littérature scientifique. Une validation géographique finale a ensuite été réalisée à l'aide des bibliothèques Cartopy et Shapely afin de vérifier la position marine des observations. Cette étape a révélé que 657 points GPS étaient localisés sur les terres émergées, principalement en raison de la résolution spatiale des grilles ICCAT ($0,5^\circ \times 0,5^\circ$). Ces observations ont été supprimées car incompatibles avec l'écologie strictement pélagique de *Thunnus thynnus*.

Enfin, un filtrage spatial complémentaire a permis d'exclure les observations situées dans la mer Noire ainsi que dans la zone atlantique du golfe de Gascogne, afin de conserver uniquement les données appartenant au bassin méditerranéen stricto sensu. Le jeu de données final comprend ainsi 32 537 observations validées de *Thunnus thynnus* (Figure 6).

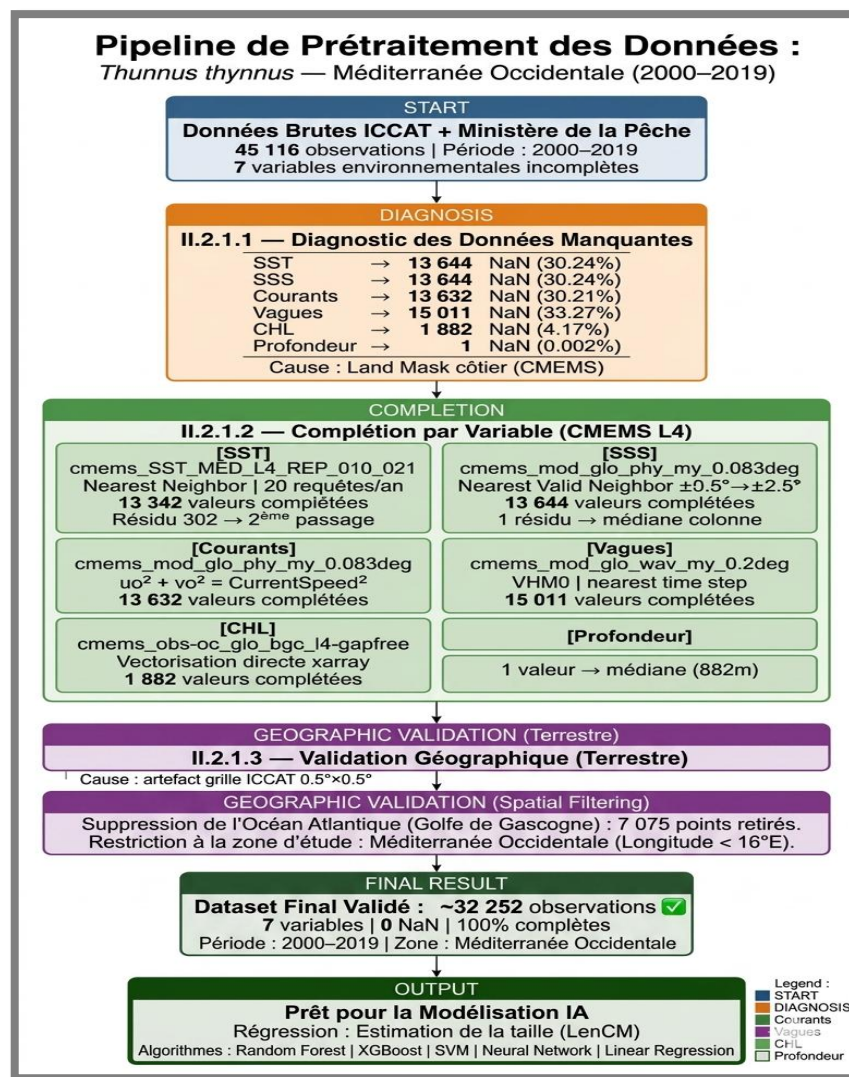


Figure 6: Schéma synthétique du pipeline Prétraitement des données (*Thunnus thynnus*), Méditerranée Occidentale (2000–2019).

3. L'analyse statistique descriptive

Constitue une étape préliminaire indispensable dans tout travail de modélisation spatio-temporelle, permettant de caractériser la structure et la distribution des données avant l'application de tout algorithme d'apprentissage automatique (Kapadia et al., 2017).

3.1 Moyenne et médiane

La moyenne représente la valeur centrale arithmétique de l'ensemble des observations, tandis que la médiane correspond à la valeur qui divise la distribution en deux parties égales. La comparaison de ces deux indicateurs permet de détecter une éventuelle asymétrie dans la distribution des données.

3.2 Écart-type et variance

L'écart-type mesure la dispersion des valeurs autour de la moyenne. Une valeur élevée indique une forte variabilité des données, ce qui est fréquemment observé pour des variables environnementales comme la SST ou la chlorophylle-a en Méditerranée.

3.3 Valeurs minimales et maximales

Les valeurs extrêmes permettent d'identifier l'étendue de chaque variable et de détecter d'éventuelles valeurs aberrantes pouvant affecter la qualité des modèles prédictifs.

3.4 Quartiles (Q1, Q2, Q3)

Les quartiles divisent la distribution en quatre parties égales. L'intervalle interquartile ($Q3 - Q1$) représente la zone de concentration de 50 % des observations et constitue un indicateur robuste de la dispersion, insensible aux valeurs extrêmes.

3.5 Asymétrie (Skewness)

Le coefficient d'asymétrie renseigne sur la forme de la distribution. Une valeur positive indique une distribution étalée vers les grandes valeurs, tandis qu'une valeur négative traduit un étalement vers les petites valeurs. Cet indicateur est essentiel pour évaluer la normalité des données avant modélisation.

3.6 Aplatissement

Le kurtosis mesure le degré de concentration des valeurs autour de la moyenne par rapport à une distribution normale. Un kurtosis élevé révèle la présence de valeurs extrêmes fréquentes, ce qui peut influencer le comportement des algorithmes d'apprentissage automatique.

3.7 Matrice de corrélation

La matrice de corrélation de Pearson et de Spearman permet de quantifier les relations linéaires et non linéaires entre les variables environnementales (SST, CHL, SSS, courants, vagues). Elle est indispensable pour identifier les redondances entre prédicteurs et orienter la sélection des variables pour la modélisation.

3.8 Histogrammes et Boxplots

Les histogrammes permettent de visualiser la distribution de fréquence de chaque variable, tandis que les boîtes à moustaches (Boxplots) représentent simultanément la

médiane, les quartiles et les valeurs aberrantes. Ces deux outils graphiques facilitent l'exploration visuelle des données avant toute modélisation.

4. Métriques et évaluation

Les modèles de régression prédisent une valeur numérique continue (ex : abondance d'espèces, CPUE, température). Pour évaluer leur qualité, on mesure l'écart entre les valeurs prédites $\hat{y}_i$ et les valeurs réellement observées y_i . Quatre métriques principales sont utilisées : MSE, RMSE, MAE et R^2 .

Le MSE mesure la moyenne des carrés des écarts entre les valeurs observées et prédites, ce qui pénalise fortement les erreurs importantes et le rend particulièrement sensible aux valeurs aberrantes. Cette métrique est largement utilisée comme fonction de perte lors de l'entraînement des modèles de régression.

Le RMSE, obtenu en prenant la racine carrée du MSE, fournit une estimation de l'erreur moyenne dans les mêmes unités que la variable cible, ce qui facilite son interprétation et explique son utilisation fréquente dans les études scientifiques.

Le MAE correspond à la moyenne des erreurs absolues et offre une mesure plus robuste en présence de valeurs extrêmes, puisque toutes les erreurs sont pondérées de manière identique.

Tableau 7: Principales métriques d'évaluation des modèles de régression.

Métrique	Nom complet	Formule mathématique	Description	Sensibilité aux valeurs aberrantes	Intervalle des valeurs	Valeur optimale	Avantages	Limites
MSE	Mean Squared Error (Erreur Quadratique Moyenne)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$	Moyenne des carrés des écarts entre valeurs observées et prédites.	Très élevée	$[0, +\infty[$	0	Différentiable; pénalise fortement les grandes erreurs; très utilisée.	Difficile à interpréter; très sensible aux valeurs aberrantes.
RMSE	Root Mean Squared Error (Racine de l'Erreur Quadratique Moyenne)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$	Racine carrée du MSE, exprimée dans l'unité originale.	Élevée	$[0, +\infty[$	0	Facile à interpréter; conserve l'unité de la variable.	Sensible aux valeurs extrêmes.
MAE	Mean Absolute Error (Erreur Absolue Moyenne)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $	Moyenne des valeurs absolues des erreurs.	Faible à modérée	$[0, +\infty[$	0	Robuste aux valeurs aberrantes; facile à interpréter.	Ne pénalise pas fortement les grandes erreurs.

R^2	Coefficient de Détermination	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	Proportion de la variance expliquée par le modèle.	Modérée	$]-\infty, 1]$	1	Évalue la qualité globale de l'ajustement.	Ne mesure pas directement l'erreur de prédiction.
-------	------------------------------	---	--	---------	----------------	---	--	---

La comparaison conjointe du MAE et du RMSE permet d'évaluer l'influence des erreurs extrêmes : lorsque leurs valeurs sont proches, les erreurs sont relativement homogènes, tandis qu'un RMSE nettement supérieur au MAE indique la présence de quelques erreurs importantes ou d'outliers. Enfin, le coefficient de détermination (R^2) quantifie la proportion de la variance observée expliquée par le modèle. Sa valeur varie généralement de 0 à 1, une valeur proche de 1 indiquant une excellente capacité explicative, tandis qu'une valeur négative traduit des performances inférieures à celles d'un modèle naïf basé sur la moyenne des observations. En pratique, aucune métrique ne permet à elle seule de caractériser pleinement la qualité d'un modèle ; il est donc recommandé d'utiliser conjointement le MAE, le RMSE et le R^2 afin d'obtenir une évaluation robuste de la précision, de la sensibilité aux erreurs extrêmes et du pouvoir explicatif du modèle.

5. Les différentes méthodes de fine-tuning

5.1 Optimisation des hyperparamètres

L'optimisation des hyperparamètres (fine-tuning) consiste à trouver la combinaison optimale de valeurs pour les paramètres de contrôle du modèle. Plusieurs techniques sont disponibles([Liam Li et al., 2020](#)) :

- **Grid Search (recherche exhaustive)** : explore systématiquement toutes les combinaisons possibles d'hyperparamètres dans un espace prédéfini. Très précis mais coûteux en temps lorsque l'espace de recherche est grand ([Dingba & Andreas, 2024](#)).
- **Random Search** : sélectionne aléatoirement des combinaisons dans l'espace des hyperparamètres. Souvent plus efficace que le Grid Search car il explore plus diversément l'espace de recherche ([J. Bergstra & Y. Bengio, 2012](#)).
- **Bayesian Optimization** : construit un modèle probabiliste (souvent un processus gaussien) de la relation entre les hyperparamètres et la performance du modèle, et l'utilise pour guider intelligemment la recherche vers les zones prometteuses ([Wu et al., 2019](#)).
- **Hyperband / ASHA** : algorithmes d'allocation adaptative des ressources qui allouent plus de temps d'entraînement aux configurations prometteuses et stoppent prématurément les mauvaises, ce qui accélère considérablement le processus ([Lisha Li et al., 2018](#)) .

Chapitre III :

Résultats Et Discussions

1. Les Résultats

Ce chapitre présente l'ensemble des résultats obtenus lors de la phase de modélisation et de prédiction de la distribution spatio-temporelle du thon rouge (*Thunnus thynnus*) en Méditerranée Occidentale. Les résultats sont organisés en deux sections principales : la première concerne target 1 (prédiction des coordonnées géographiques LonX et LatY), et la seconde target 2 (prédiction de la longueur LenCM). Pour chaque target les performances des modèles sont évaluées avant et après optimisation des hyperparamètres par Bayesian Optimisation et Grid Search, suivies d'une analyse cartographique détaillée des prédictions sur les données Copernicus (2020–2026).

1.1. Prédiction de position de distribution de thon rouge (Target 1)

Target 1 vise à prédire simultanément la longitude (LonX) et la latitude (LatY) du thon rouge à partir des variables environnementales satellitaires. Les métriques utilisées — R^2 , RMSE et MAE — sont calculées de manière globale sur l'ensemble des coordonnées (LonX et LatY), conformément à la correction méthodologique appliquée. La distance d'erreur est exprimée en kilomètres via la formule de Haversine, qui tient compte de la courbure terrestre.

1.1.1. Performances des modèles avant et après optimisation des hyperparamètres

Afin d'évaluer l'apport de l'optimisation des hyperparamètres sur les performances prédictives des modèles, une comparaison systématique a été réalisée entre les modèles bruts (avant fine-tuning) et les modèles optimisés (après recherche bayésienne ou Grid Search). Le tableau 9 présente les résultats obtenus pour la prédiction de la position géographique (Target 1) du thon rouge en Méditerranée occidentale, à travers trois métriques d'évaluation : le coefficient de détermination (R^2), la racine de l'erreur quadratique moyenne (RMSE) et l'erreur absolue moyenne (MAE).

Chapitre III : Résultats Et Discussions

Tableau 8: Performances des modèles avant et après optimisation des hyperparamètres pour la prédiction de la position géographique du thon rouge (target 1)

Modèle	R ² (avant)	R ² (après)	RMSE (avant) (ind/pixel)	RMSE (après) (ind/pixel)	MAE (avant) (ind/pixel)	MAE (après) (ind/pixel)
XGBoost	0,9716	0,9777	1.13	0.79	0.07	0.06
Random Forest	0,9751	0,9752	0.31	0.32	0.03	0.03
SVR	0,9136	0,9328	1.96	1.96	0.09	0.09
Linear Regression	0,6817	0,6817	2.02	2.02	0.11	0.11

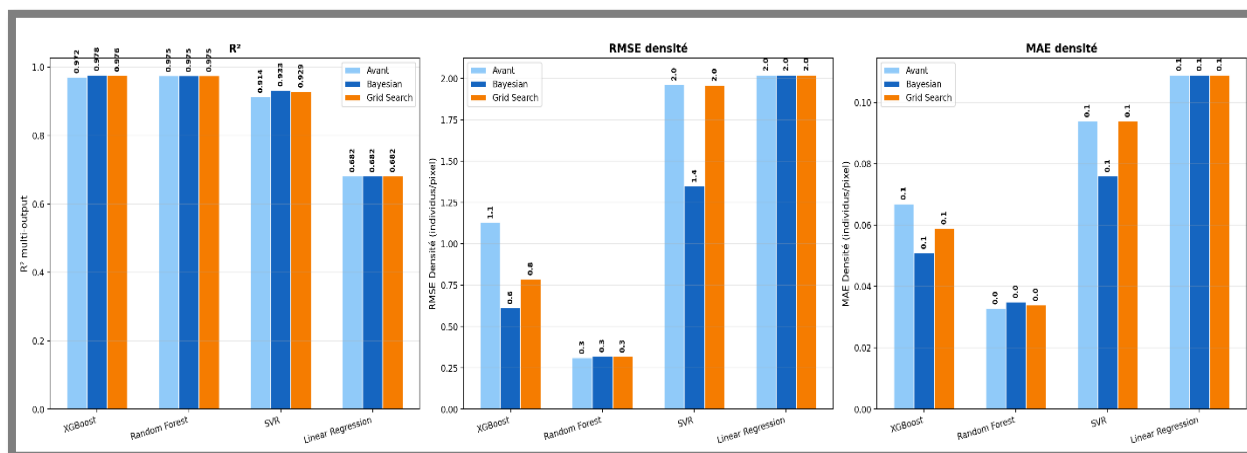


Figure 7: comparative des métriques (R², RMSE, MAE) pour les quatre modèles, avant optimisation et après optimisation par recherche bayésienne et Grid Search

Le tableau 10 et la figure 7 présente les performances comparatives de quatre modèles d'apprentissage automatique pour la prédiction la position du thon rouge en Méditerranée occidentale, avant et après optimisation des hyperparamètres.

Avant fine-tuning, le modèle Random Forest donne les meilleures performances avec un coefficient de détermination le plus élevé (**R² = 0,9751**) ainsi que les plus faibles erreurs (RMSE = 0.31 ind/pixel et MAE = 0.03 ind/pixel), indiquant une excellente capacité à reproduire la variabilité spatiale observée de l'espèce. Le modèle XGBoost présente également des performances très élevées avec un **R² de 0,9716**, bien que ses erreurs soient

Chapitre III : Résultats Et Discussions

légèrement supérieures (RMSE =1.13 ind/pixel), traduisant une moindre précision dans la localisation des individus. Cette supériorité des modèles Random Forest et XGBoost sur les autres approches suggère que les relations entre les variables environnementales et la position géographique du thon rouge sont fortement non linéaires. Le SVR affiche des performances intermédiaires ($R^2 = 0,9136$; RMSE = 1.96 ind/pixel ; MAE =0.09 ind/pixel), tandis que le modèle le plus faible est la Régression Linéaire avec un R^2 de 0,6817, un RMSE de 2.02 ind/pixel et une MAE de 0.11 ind/pixel, ce qui confirme que les relations entre les variables explicatives et la variable expliquée ne peuvent être correctement représentées par un modèle linéaire simple.

Après fine-tuning, le modèle XGBoost amélioré par grid search devient le modèle le plus performant avec un R^2 de **0,9777**. Le Random Forest conserve presque les mêmes performances avant fine tuning ($R^2 = 0,9752$; RMSE =0.32), confirmant sa robustesse naturelle face aux hyperparamètres. Le SVR amélioré avec une augmentation légère de R^2 ($R^2 = 0,9328$) cette augmentation légère confirme l'optimisation du modèle. Au contraire à la Régression Linéaire garde les mêmes performances après optimisation par grid search, ses limites étant structurelles et non liées aux hyperparamètres.

Ces résultats confirment la capacité des modèles à capturer les relations complexes et non linéaire entre les variables et la distribution du thon rouge.

Dans l'ensemble les résultats basés sur l'intelligence artificielle en particulier des méthodes optimisées, forment des outils performants pour la modélisation de la distribution spatio-temporelle de thon rouge dans le bassin méditerranéen

1.1.2. Analyse spatiale

1.1.2.1. Résultat de test

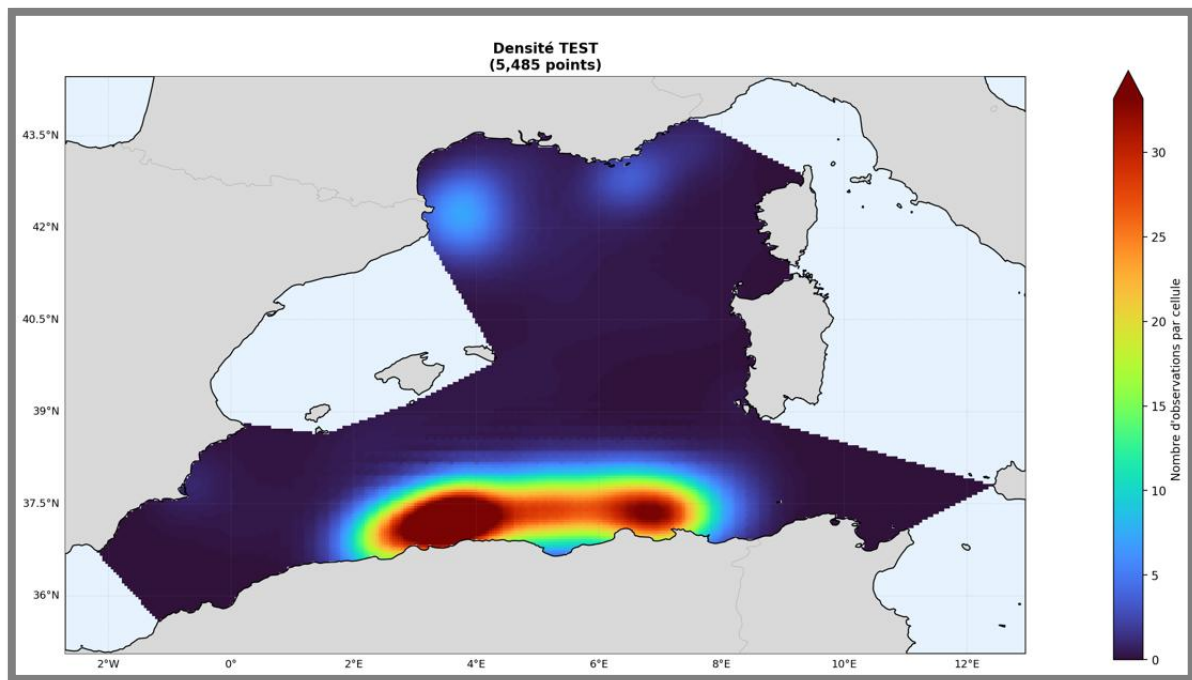


Figure 8: Densité spatiale du Thon Rouge (KDE — Ordinary Kriging) — Test set (Bassin Occidental)

La figure 8 présente la densité spatiale du thon rouge obtenue après le test de model XGBoost prédite optimisé par Grid Search sur le jeu de données de test (bassin occidental), obtenue à partir du modèle XGBoost (target 1).

Le long des côtes italiennes française espagnole et tunisienne la densité é spatiale est presque faible et même nulle au nord de bassin on remarque deux noyaux du couleur bleu clair qui reflète à une densité de 5 à 10 individus par pixel.

Au sud de bassin occidentale ; le long des côtes algériennes plus précisément dans le bassin algérien on remarque une forte densité avec une couleur qui varie de bleu clair au rouge foncée qui correspond a nombre d'individu supérieur à 10 individus par pixel. et même supérieur à 30.

La zone de forte densité de thon rouge constitue un milieu écologique favorable, qui peut être associée à une zone de reproduction, à une route de migration, ou à une zone favorable en raison de facteurs environnementaux tels que la température et la chlorophylle (CHL).

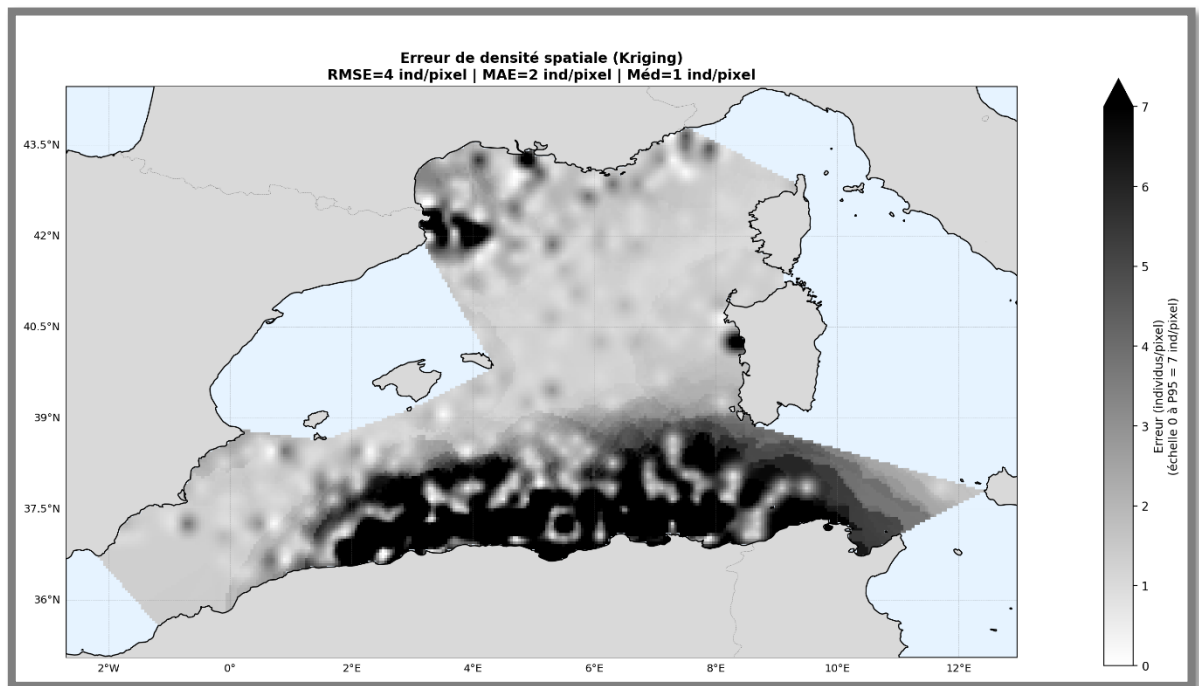


Figure 9: Erreur du modèle (Ordinary Kriging) — Test set (Bassin Occidental) | Pixels 0.045° | Palette blanc-noir | XGBoost (target 1)

La figure 9 présente la carte de répartition spatiale de erreurs absolu entre les densité observée et les densité prédite par le modele du modèle après le test du modèle XGBoost (target 1) optimise par grid search dans la zone de bassin méditerrané occidentale, les indicateur globaux sont de MAE=2 ind/pixel RMSE= 4 ind/pixel et LA MEDI =1 ind/pixel

La majorité de la superficie du bassin présent des erreurs faibles a modérée (inferieure a 2 individus par pixel) cela explique que le model est robuste dans la majorité du bassin.

Les plus fortes erreurs localisées au sud de bassin principales dans les zones de forte densité prédite cette situation sont attendus aux géostatistiques :

Les zones présentant les densités les plus élevée sont celle des zones au la variabilité est plus élevée, le model tend à produire des erreurs plus importantes dans des zones biologiquement plus actives.

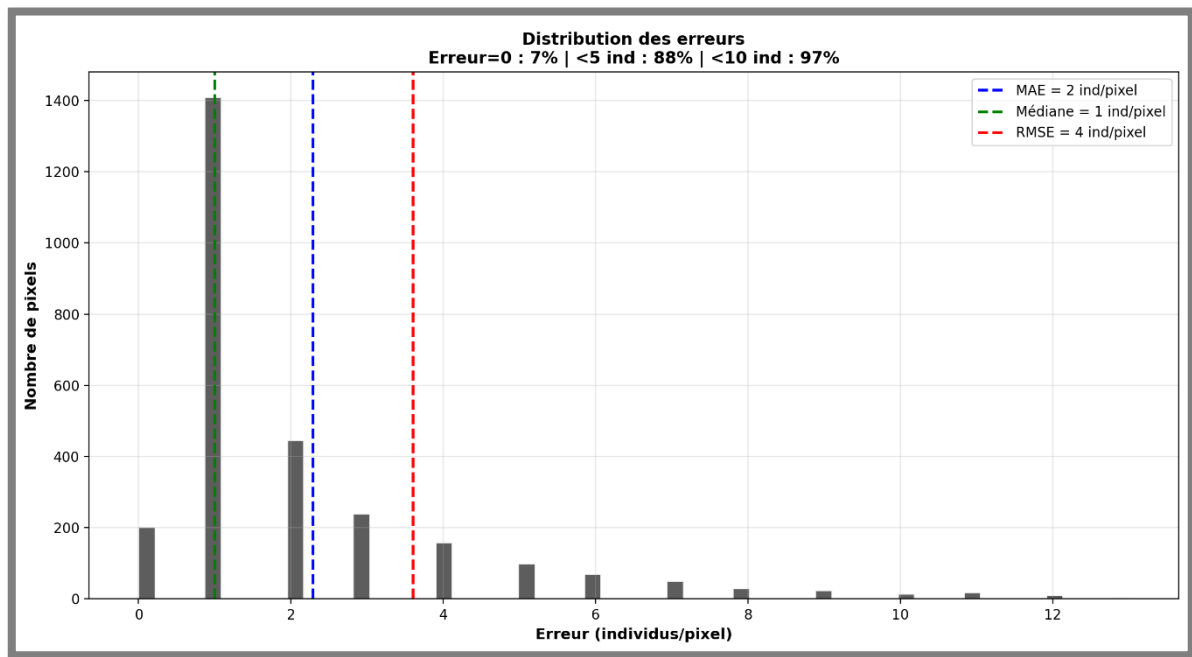


Figure 10: Histogramme des distances d'erreur géographique (individus/pixel) du modèle XGBoost (target 1) avec indication des métriques MAE, médiane et RMSE

La figure 10 présente l'histogramme de fréquences des erreurs absolue exprime en individu par pixel. L'histogramme montrent que 7 % des pixels représente une erreur nulle, 88 des pixels ont une erreur inférieure à 5 individus par pixel et 97 % des pixel affiche une erreur inférieure a 10 individus par pixel ; indiquant une majorité de prédiction précise représente la distribution est fortement asymétrique vers les faibles erreurs car ces valeurs sont inférieure a la valeur de MAE (12 individus par pixel)

La robustesse du modèle est témoignée par la majorité des erreurs les plus faible le long de bassin occidentale

CHAPITRE III : Résultats Et Discussions

1.2. Prédiction de la Longueur du Thon Rouge (Targte 2)

Target 2 vise à prédire la longueur individuelle (LenCM, en centimètres) du thon rouge à partir des variables environnementales satellitaires disponibles dans la base Copernicus. Il convient de rappeler que LenCM est une variable biologique absente de la base de prédiction Copernicus et que le modèle a été entraîné exclusivement sur les variables environnementales, ce qui rend un R^2 supérieur à 0.80 particulièrement satisfaisant.

1.2.1. Performances des modèles avant et après optimisation des hyperparamètres

Contrairement à la prédiction de la position géographique, la variable target LenCM (longueur du thon rouge) constitue une donnée biologique absente des bases de données environnementales Copernicus. Le modèle a donc été entraîné exclusivement sur les variables environnementales (température, salinité, bathymétrie, etc.). Dans ce contexte, un R^2 supérieur à 0,80 représente une performance particulièrement satisfaisante. Le tableau 9 compare les résultats des quatre modèles avant et après optimisation des hyperparamètres pour la prédiction de la longueur (target 2), à travers les métriques R^2 , MAE (cm) et RMSE (cm).

Tableau 9: Performances des modèles avant et après optimisation des hyperparamètres pour la prédiction de la longueur du thon rouge (target 2 — LenCM)

Modèle	R^2 (avant)	R^2 (après)	MAE (avant)	MAE (après)	RMSE (avant)	RMSE (après)
XGBoost	0,8002	0,8060	17,39 cm	17,22 cm	27,44 cm	27,03 cm
Random Forest	0,8023	0,8095	17,05 cm	16,89 cm	27,29 cm	26,79 cm
SVR	0,7635	0,7688	19,40 cm	18,93 cm	29,85 cm	29,52 cm
Linear Regression	0,6470	0,6470	26,80 cm	26,80 cm	36,47 cm	36,47 cm

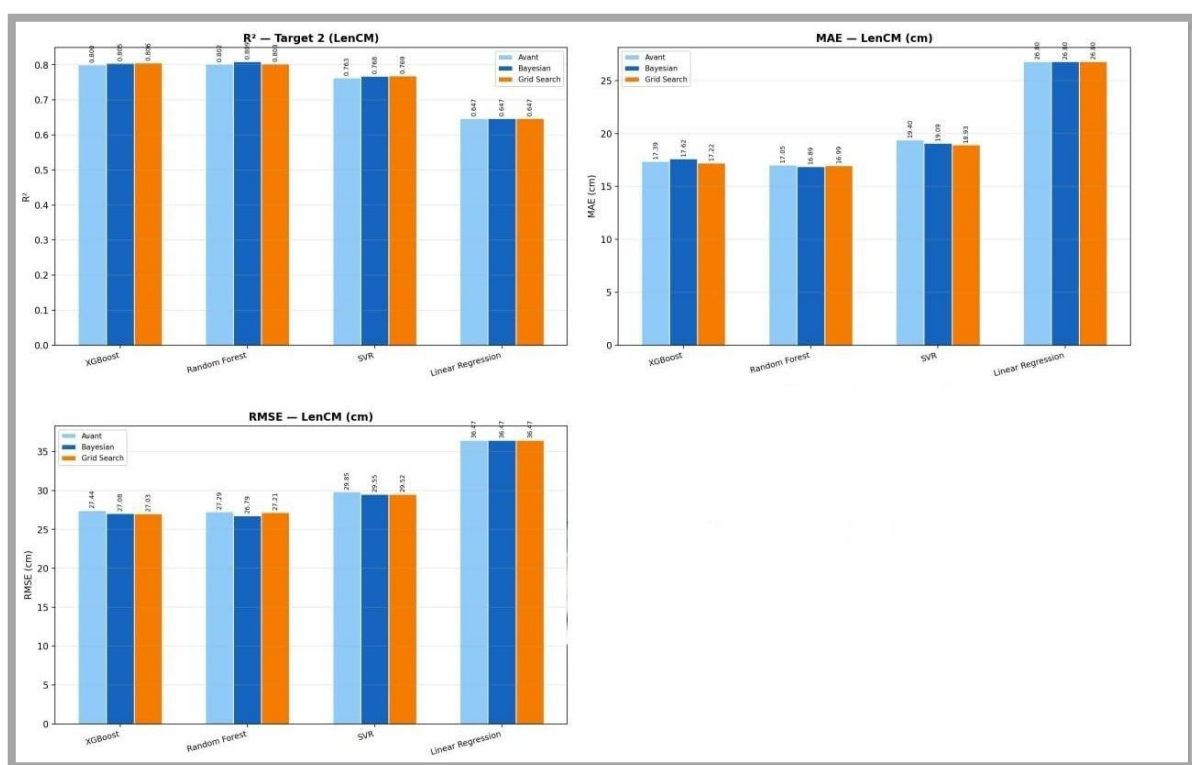


Figure 11: Comparative des métriques (R^2 , RMSE, MAE) pour les quatre modèles, avant optimisation et après optimisation par recherche bayésienne de target 2 (Grid Search pour la régression linéaire)

Le tableau 11 et la figure 9 comparent les performances de quatre modèles d'apprentissage automatique pour la prédiction de la longueur du thon rouge (target 2 - LenCM) en Méditerranée occidentale, avant et après optimisation des hyperparamètres.

Avant fine-tuning, le modèle Random Forest donne les meilleures performances avec un R^2 de 0,8023 ainsi que les plus faibles erreurs (**RMSE = 27,29 cm et MAE = 17,05 cm**), indiquant une forte capacité à estimer la taille des individus à partir des variables environnementales. Le modèle XGBoost présente des performances très proches avec un R^2 de 0,8002, bien que ses erreurs soient légèrement supérieures (**RMSE = 27,44 cm**). Le SVR affiche des performances intermédiaires (**$R^2 = 0,7635$; RMSE = 29,85 cm ; MAE = 19,40 cm**), tandis que le modèle le plus faible est la Régression Linéaire avec un R^2 de 0,6470, un RMSE de 36,47 cm et une MAE de 26,80 cm, confirmant son inadaptation à capturer les relations complexes entre les facteurs environnementaux et la taille du thon rouge.

Après fine-tuning par recherche bayésienne, le modèle Random Forest conserve la première position avec des performances améliorées (**$R^2 = 0,8095$; MAE = 16,89 cm ; RMSE = 26,79 cm**), suivi du XGBoost (**$R^2 = 0,8060$; MAE = 17,22 cm**). Le SVR occupe la position

Chapitre III : Résultats Et Discussions

intermédiaire ($R^2 = 0,7688$; MAE = 18,93 cm ; RMSE = 29,52 cm), et la Régression Linéaire reste en dernière position avec des performances inchangées ($R^2 = 0,6470$; MAE = 26,80 cm ; RMSE = 36,47 cm).

1.2.2. Analyse Spatiale

1.2.2.1. Résultats de test

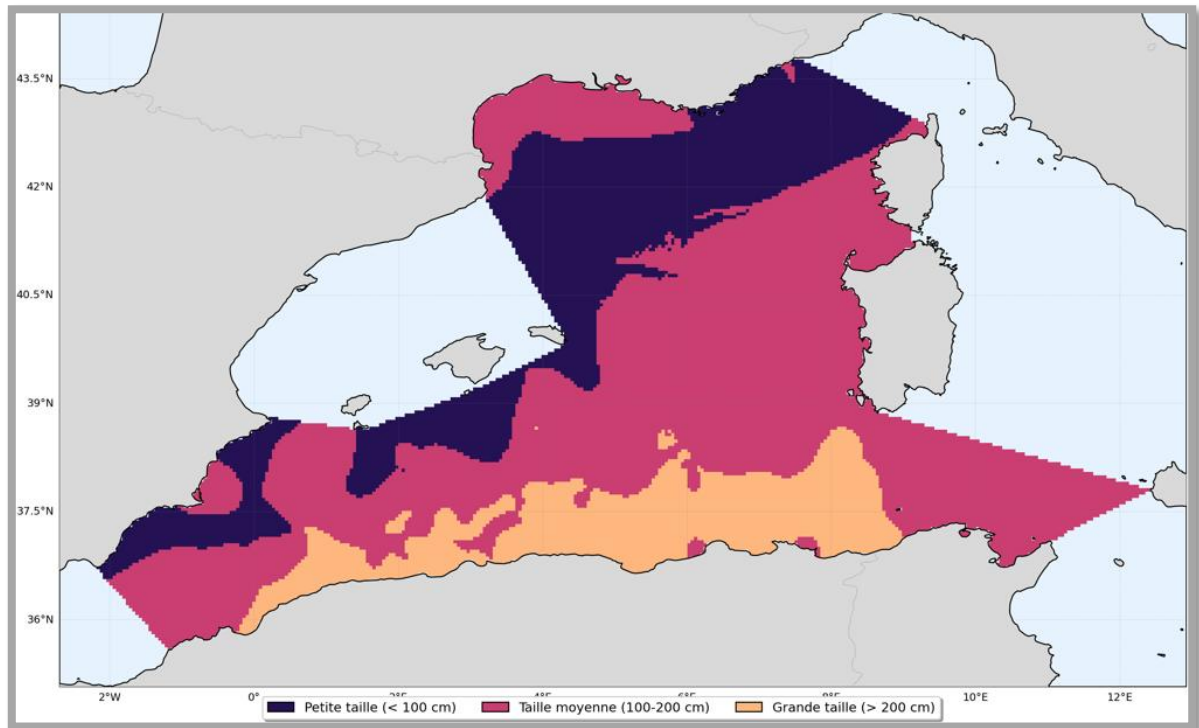


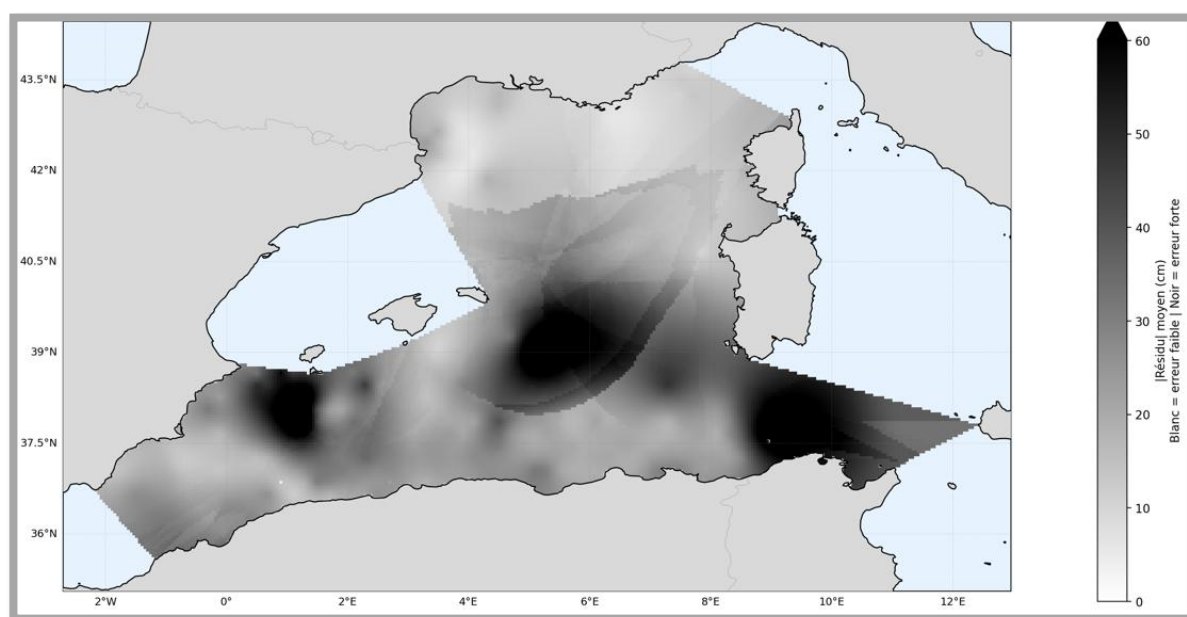
Figure 12: Distribution spatiale des classes de taille du thon rouge (petite, moyenne, grande taille) en Méditerranée occidentale — Test set — Random Forest Bayesian (target 2)

La figure 12 présente la distribution spatiale des classes de taille du thon rouge (petite, moyenne et grande) en Méditerranée occidentale, obtenue à partir du modèle Random Forest optimisé par recherche bayésienne (target 2 - LenCM). La carte révèle une ségrégation spatiale marquée selon un gradient nord-sud : les individus de petite taille (moins de 100 cm) se concentrent dans le nord du bassin, les individus de grande taille (plus de 200 cm) dominent le sud, tandis que les individus de taille moyenne (100-200 cm) occupent une zone de transition entre ces deux régions ainsi que le long des côtes françaises et tunisiennes.

Cette organisation spatiale est cohérente avec plusieurs processus écologiques connus chez les grands pélagiques. La présence des individus de petite taille dans le nord du bassin suggère l'existence de zones de nurserie, tandis que les individus de grande taille au sud

Chapitre III : Résultats Et Discussions

reflètent une préférence pour des habitats associés à des conditions océanographiques spécifiques (température, salinité, courants). La dominance des individus de taille moyenne dans le centre du bassin indique des conditions environnementales favorables à la croissance de l'espèce. Le modèle de prédiction s'avère robuste pour les classes petite et moyenne, mais moins performant pour les individus de taille supérieure à 200 cm. Cette distribution propose une ségrégation écologique des individus selon leur taille, renforçant l'hypothèse d'une utilisation différenciée de l'habitat en fonction des stades de développement.



La figure 13 présente la carte de distribution spatiale de la magnitude des résidus absolu du test du modèle Random Forest optimisé par recherche bayésienne la prédiction de la longueur du thon rouge (target 2 - LenCM) sur le jeu de données de test en Méditerranée occidentale, avec une MAE de 21,94 cm et un RMSE de 31,76 cm. Les résidus sont exprimés en centimètres, allant du blanc (erreurs faibles) au noir (erreurs élevées).

La figure représente une forte hétérogénéité spatiale des erreurs, On remarque deux types de zone forte erreur qui couvre la zone centrale de bassin sud West et sud-est de bassin dans ces zones les résidus dépassent 50 cm

Donc faibles erreurs couvrent le nord et sud du bassin ces zones présente des résidus inférieur a 20 cm donc ces régions les conditions environnementales peut être correctement près en considération par les variables explicatives utilisée pour le modèle en comparant la figure la

figure 12 et 13 conclu que le modèle est faible dans la prédiction des individus de grand de taille et forte dans la prédiction des individus moyenne et petites

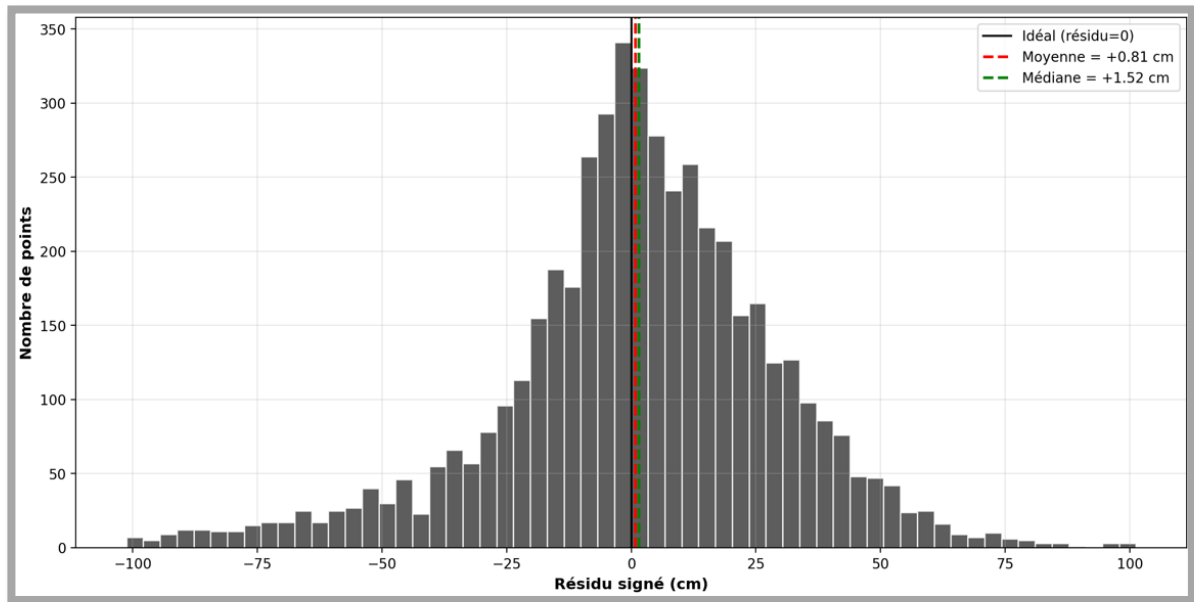


Figure 14: Histogramme des résidus signés — Test set Méditerranée occidentale — Random Forest Bayesian (target 2)

La figure 14 présente l'histogramme des résidus absolu du modèle Random Forest optimisé par recherche bayésienne pour la prédiction de la longueur du thon rouge (target LenCM) sur le jeu de données de test en Méditerranée occidentale. L'histogramme montre une distribution symétrique au tour de la moyenne de 0,81 cm, la distribution suit la loi normale cela explique une distribution parfaite entre les valeurs surestimées et sous-estimées. Cette distribution est un bon indicateur de la fiabilité du modèle.

1.3. Analyse Globale et Importance des Variables

Cette section analyse l'importance relative des variables environnementales dans les prédictions des modèles finaux, afin d'identifier les facteurs clés régissant la position géographique et la longueur du thon rouge en Méditerranée occidentale.

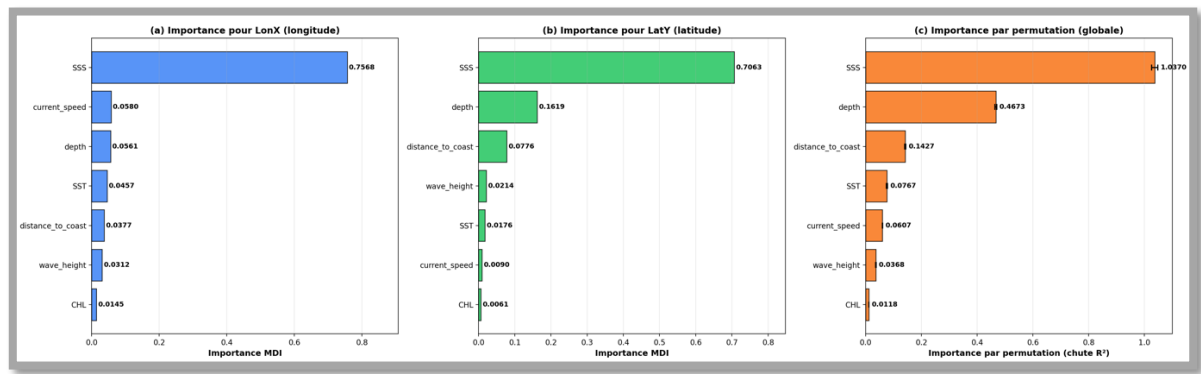


Figure 15: Importance des variables pour target 1 (position LonX + LatY) — Modèle XGBoost + MultiOutputRegressor + Grid Search — Méditerranée

Tableau 10: Tableau récapitulatif du classement combiné des variables pour target 1 (Position) — Modèle XGBoost + Grid Search — Méditerranée

Feature	MDI LonX	MDI LatY	MDI moy
SSS	0.7568	0.7063	0.7315
depth	0.0561	0.1619	0.1090
distance_to_coast	0.0377	0.0776	0.0577
SST	0.0457	0.0176	0.0316
current_speed	0.0580	0.0090	0.0335
wave_height	0.0312	0.0214	0.0263
CHL	0.0145	0.0061	0.0103

La figure 15 et le tableau 11 présentent l’importance des variables environnementaux pour la prédiction de la position géographique du thon rouge (target 1), obtenue à partir du modèle XGBoost optimisé par Grid Search,

Les graphiques et le tableau montrent que la première variable la plus important pour la prédiction est **la Salinité de Surface (SSS)** avec **MDI de 0, 7568**. L’importance de cette variable est dû que ce dernier est une caractéristique hydrologique.

Chapitre III : Résultats Et Discussions

La **bathymétrie (depth)** est importante pour la prédiction en deuxième position avec des scores **MDI = 0.1090**, indiquant que la profondeur joue un rôle important dans la répartition de l'espèce, les thons préfèrent les eaux profondes du large aux zones côtières peu profondes.

La **température** et la **distance à la côte** représente les variables les plus important de prédiction en 3 positions avec des score **MDI= 0.0316** ET **MDI=0.0577**

La **vitesse des courants**, est le quatrième paramètre important de la prédiction avec des scores **MDI=0.0335**

La **hauteur des vagues** et le cinquième paramètre important de la prédiction avec des scores de **MDI=0.0103**

Enfin la **chlorophylle-a (CHL)** en dernier position avec des scores de **MDI=0 ,0103**

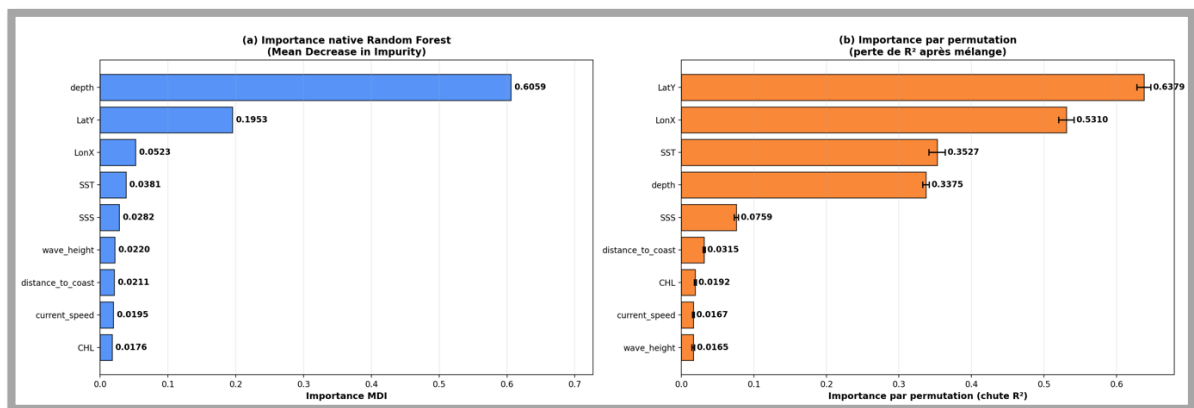


Figure 16: Importance des variables (MDI et permutation) pour target 2 (LenCM) — Modèle Random Forest Bayesian— Méditerranée

Chapitre III : Résultats Et Discussions

Tableau 11: Tableau récapitulatif du classement combiné des variables pour target 2 (Longueur en cm) — Modèle Random Forest Bayesian — Méditerranée

Feature	MDI	Rank MDI	Permutation
depth	0,6059	1	0,3375 ± 0,0045
SST	0,0381	2	0,3527 ± 0,0113
SSS	0.0282	3	0.0759 ± 0.0029
distance_to_coast	0,0211	4	0,0759 ± 0,0029
wave_height	0,0211	5	0,0315 ± 0,0018
current_speed	0,0195	6	0,0167 ± 0,0011
CHL	0,0176	7	0,0192 ± 0,0013

La figure 16 et le tableau 12 présentent l'importance des variables environnementales pour la prédiction de la longueur du thon rouge (target 2), obtenue à partir du modèle Random Forest optimisé par La recherche Bayesian.

La figure 16 et le tableau 12 montrent des variables explicatives selon leur contribution globale à la prédiction d'après **MDI**. On a 3 catégories de classement la première catégorie représente **la profondeur** avec une forte contribution **MDI= 0,6059**

La 2 catégories représente 3 variables avec une contribution moyenne a la prédiction comme suite **SST**, **SSS**, **distance à la coute** avec un **MDI=0,0381**, **MDI=0.0282**, **MDI=0,0211** respectivement

Et enfin la 3 catégories qui représente 3 variables **la hauteur de la vague**, **La vitesse de courant** et **chlorophylle** avec une contribution faible **MDI=0,0211**, **MDI=0,0195**, **MDI=0,0176** respectivement

1.4 Résultat de la corrélation observe vs prédit

Cette partie présente les résultats spatiaux temporelle de comparaison des cartes observée avec celle prédite pour target 1 et target 2

- **Target 1 : la position géographique**

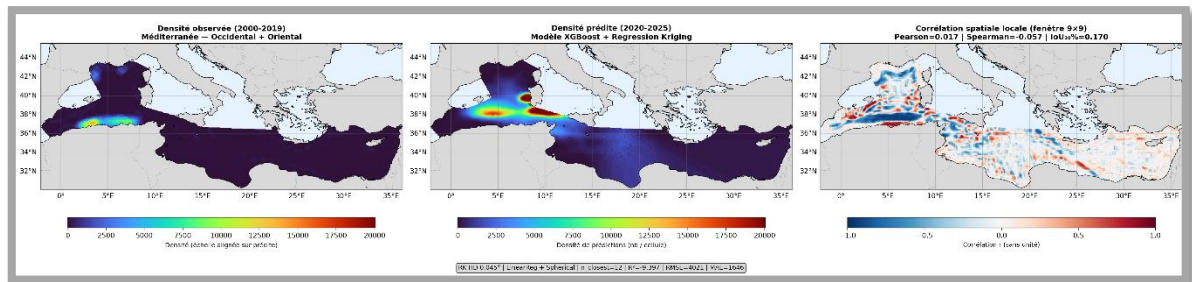


Figure 17: corrélation spatiale de la densité observée et la densité prédite par le modèle xg boost dans le bassin Méditerranée — Densité observée (2000–2019) vs Densité prédite (2020–2025)

La figure 17 illustre la corrélation spatiale entre la densité observée du thon rouge et la densité prédite par le modèle dans le bassin méditerranéen.

En comparant les résultats de la carte observée et celle prédite on remarque que les fortes densités se concentrent presque dans les mêmes zones de bassin occidentale. Cette coïncidence confirme que la capacité du modèle à identifier les habitats de l'espèce avec une prédiction très forte.

Les métriques de la carte de corrélation spatiale (Pearson = 0,017 ; Spearman = -0,057 ; IoU = 0,170) montrent une faible corrélation globale entre la carte de prédiction et la carte observée dans le bassin méditerranéen. Toutefois, la valeur $\text{IoU}_{20\%} = 0,170$ indique une superposition partielle des zones de forte densité suggérant que le modèle identifie correctement les zones de forte densité.

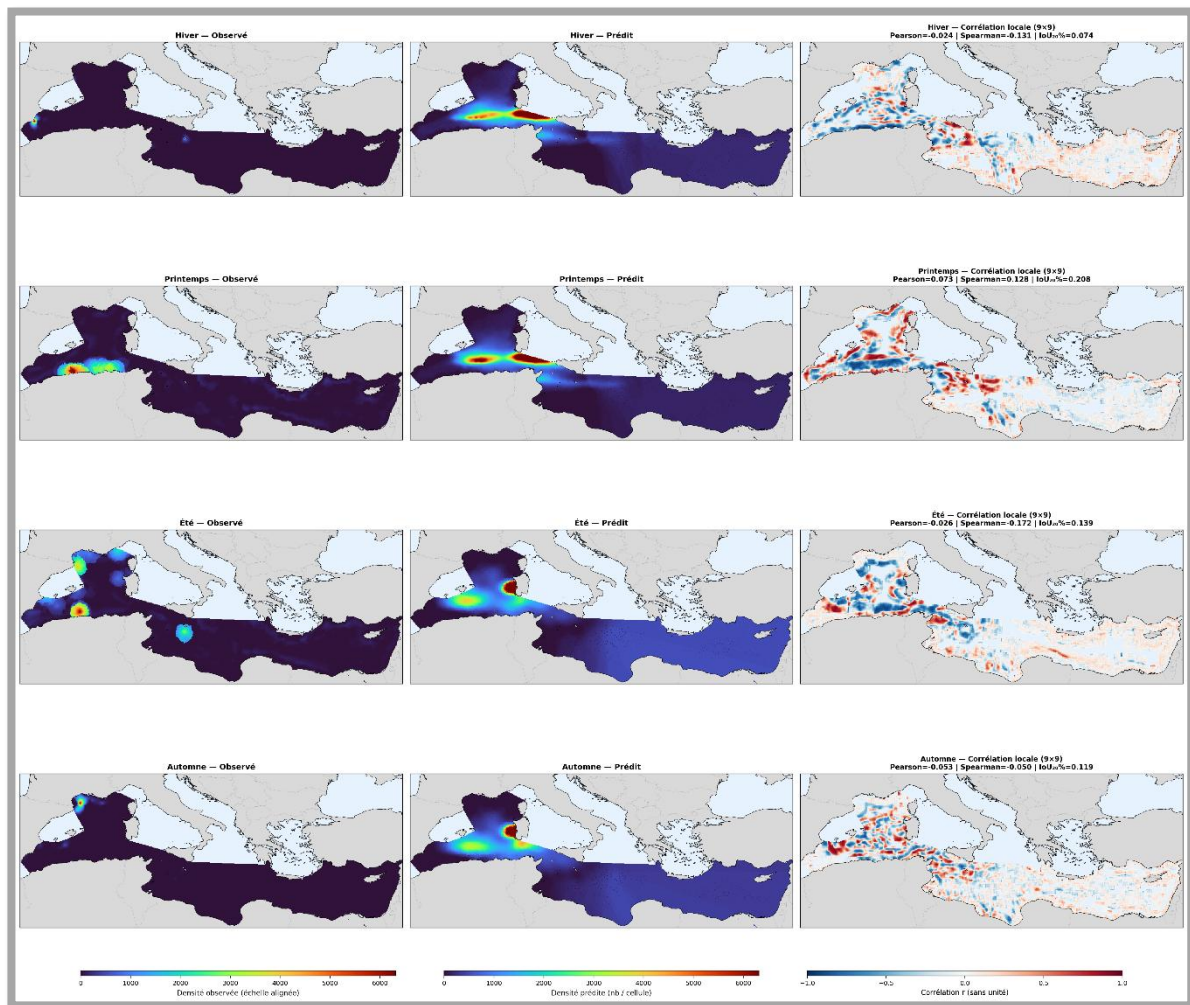


Figure 18: corrélation saisonnière de la densité spatiale du thon rouge dans le bassin Méditerranée — Densité observée vs Densité prédite par le modèle xg boost

La figure 18 présente la comparaison saisonnière de la densité spatiale du thon rouge entre les observations historiques (2000-2019) et les prédictions du modèle XGBoost optimisé par Grid Search avec interpolation par Regression Kriging (2020-2025), sur l'ensemble du bassin méditerranéen. Les cartes d'observation montrent une dynamique saisonnière marquée :

L'Été concentre les plus fortes densités dans le bassin occidental, notamment dans le golfe du Lion, au large des Baléares et dans le détroit de Sicile-Tunisie, correspondant à la période de reproduction de l'espèce en Méditerranée. Le Printemps et l'Automne présentent des densités intermédiaires, reflétant les phases de migration, tandis que l'Hiver affiche les densités les plus faibles avec une dispersion des individus vers l'Atlantique. Les cartes prédites reproduisent exactement cette même dynamique saisonnière, avec un pic estival marqué dans les mêmes zones de concentration, confirmant la capacité du modèle à identifier la route migratoire connue de *Thunnus thynnus*. Les cartes prédites une

Chapitre III : Résultats Et Discussions

distribution plus étalée régulière que l'observée. Le modèle prédit bien les zones d'abondance, mais il surestime sous-estime certaines zones selon la saison. Cela signifie que le modèle a pris en considération les grandes tendances spatiales, mais qu'il manque une partie du signal écologique précis. Les cartes de corrélation montrent des valeurs faibles à modéré

Hiver (Pearson = -0,024 ; Spearman = -0,131 ; IoU_{20%} = 0,074)

Printemps (Pearson = 0,073 ; Spearman = 0,128 ; IoU_{20%} = 0,208)

Été (Pearson = -0,026 ; Spearman = -0,173 ; IoU_{20%} = 0,139)

Automne (Pearson = -0,033 ; Spearman = -0,050 ; IoU_{20%} = 0,119)

Selon ces valeurs la relation linéaire entre observée et prédite est très faible surtout en hiver le Spearman est parfois un peu meilleur que Pearson ce qui suggère que le modèle estime mieux en relation non linéaire que linéaire, la performance locale reste limitée les fortes corrélations locales au globale indiquent une capacité prédictive parfaite et les faibles corrélations indiquent une capacité prédictive imparfaite, Ces résultats ne signifient pas que le modèle est mauvais. Ils indiquent plutôt que le modèle est utile pour réaliser des prédictions à l'échelle globale, mais qu'il nécessite encore certaines améliorations.

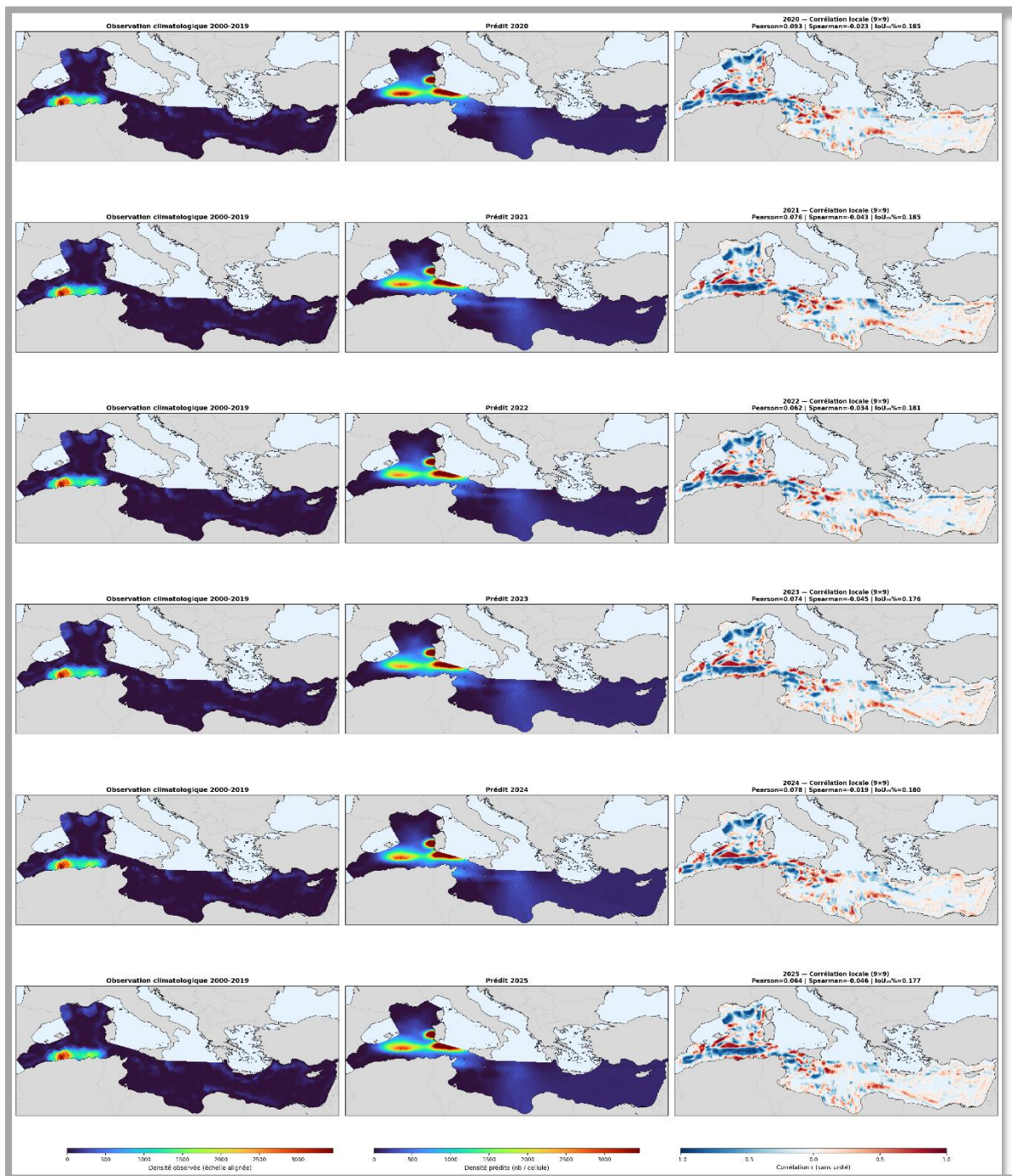


Figure 19: corrélation annuelle de la densité spatiale temporelle du thon rouge dans le bassin Méditerranéen — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle XGBoost

La figure 19 montre la distribution spatio-temporelle du thon rouge observée et prédite avec le modèle XGBoost, dans l'ensemble le bassin méditerranéen.

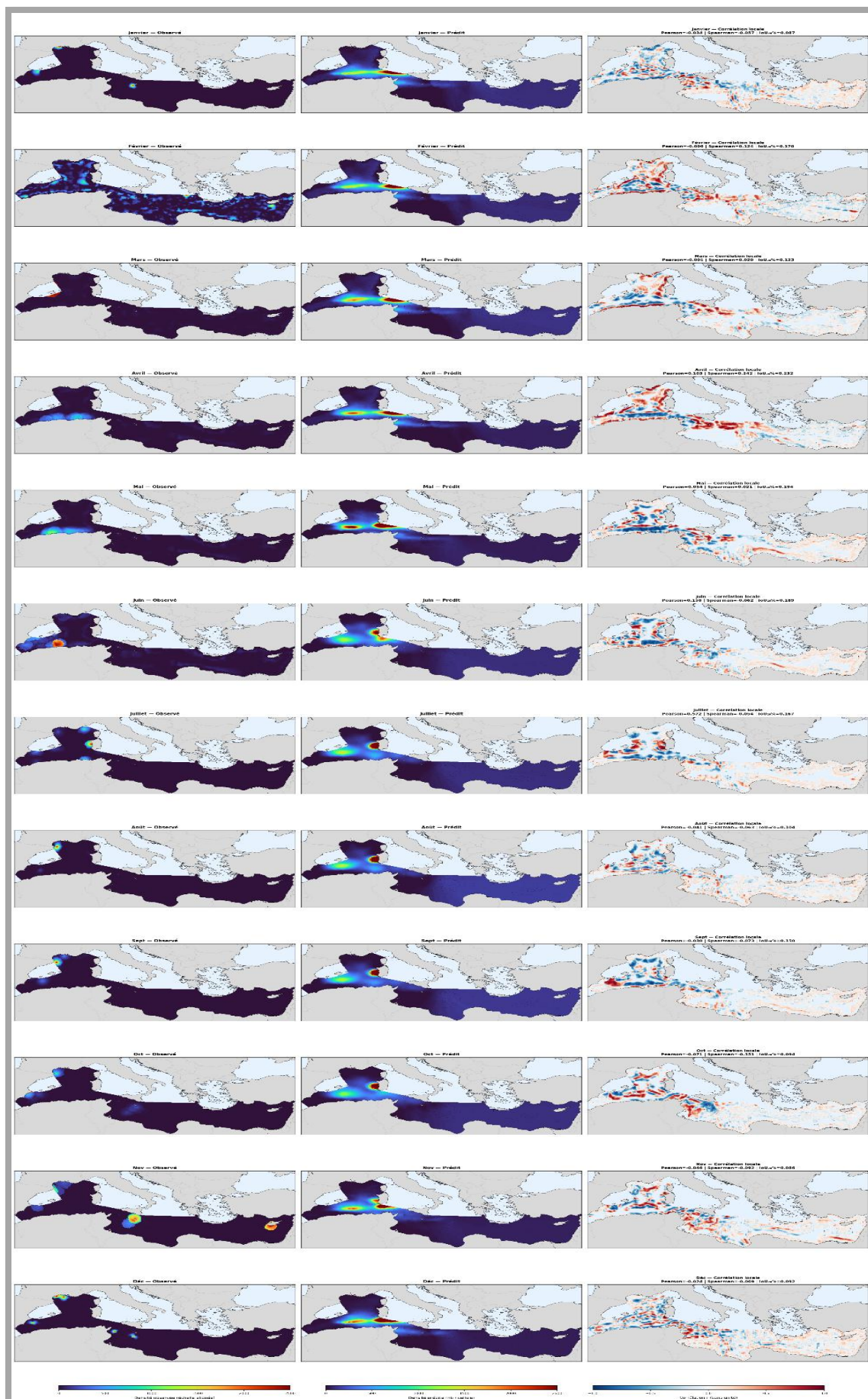
Carte d'observation, La distribution spatiale de l'espèce remarquable avec une abondance importante dans le bassin occidental notamment dans les côtes françaises le bassin algérien et même les côtes espagnole et tunisienne le modèle estime exactement dans les mêmes zones

Chapitre III : Résultats Et Discussions

de forte concentration dans la carte d'observation cette prédiction témoignée la robustesse du modèle dans la prédiction en associant les variables environnementales

Carte de prédiction annuelle les six cartes annuelle de montre une stabilité de la distribution du thon dans le bassin méditerranée le modèle prédit exactement les même zones de forte concentration de thon rouge dans le bassin occidentale sur ensemble de la période , ca forte liaison non linéaire entre les variables environnementaux et l'habitat de l'espèce on observe une légère tendance l'intensification dans certaine zone 2023-2025 qui pourrait explique par une variation environnementaux interannuelle

L'analyse spatio-temporelle des corrélations met en évidence la présence de corrélations positives, négatives ainsi que de corrélations faibles à l'échelle de l'ensemble du bassin méditerranéen, en particulier dans le bassin occidental. Cette variabilité des corrélations peut s'expliquer par une surestimation ou une sous-estimation des prédictions du modèle dans certaines zones. Par ailleurs, la persistance de faibles corrélations dans le bassin oriental souligne la nécessité d'intégrer des données couvrant l'ensemble du bassin méditerranéen afin d'améliorer la capacité de généralisation spatiale du modèle.



Chapitre III : Résultats Et Discussions

Figure 20: corrélation mensuelle mensuelle de la densité spatiale prédite vs observée du thon rouge dans le bassin méditerranéen — par le modèle XGBoost

La figure 20 présente l'évolution mensuelle de la densité spatiale prédite du thon rouge dans le bassin méditerranéen par le modèle XGBoost.

La carte observée mensuelle montre une dynamique saisonnière marquée.

La densité augmente progressivement à partir du mois de mars pour atteindre un pic maximal en juin et juillet. Cette période correspond à la saison de reproduction de l'espèce dans le bassin méditerranéen. Elle diminue ensuite à partir du mois de septembre jusqu'en février. Les mois d'été (juin, juillet, août) concentrent les plus fortes densités dans le bassin occidental. Ces zones de forte abondance sont connues comme des habitats préférentiels pour la reproduction, l'alimentation et les routes de migration du thon rouge. Les mois d'hiver (décembre, janvier, février) présentent une densité plus faible, avec une distribution des individus vers l'Atlantique, reflétant la phase de migration post-reproductive vers l'Atlantique.

Ce cycle mensuel démontre la capacité du modèle à intégrer la variabilité temporelle fine de la distribution du thon rouge, que ce soit pour la reproduction, la migration ou l'alimentation. Ces résultats confirment également l'apport de l'intelligence artificielle pour le suivi spatiotemporel du thon rouge dans la zone d'étude. Il reste toutefois à évaluer la corrélation avec les données observées.

Les cartes de corrélation montrent des valeurs de corrélation faibles à modérées selon les mois : Janvier (Pearson = -0,034 ; IoU = 0,087), Février (Pearson = -0,006 ; IoU = 0,170), Mars (Pearson = -0,001 ; IoU = 0,123), Avril (Pearson = 0,103 ; IoU = 0,232), Mai (Pearson = 0,054 ; IoU = 0,194), Juin (Pearson = 0,158 ; IoU = 0,189), Juillet (Pearson = 0,572 ; IoU = 0,167), Août (Pearson = -0,041 ; IoU = 0,104), Septembre (Pearson = -0,030 ; IoU = 0,150), Octobre (Pearson = -0,071 ; IoU = 0,094), Novembre (Pearson = -0,046 ; IoU = 0,086), Décembre (Pearson = -0,024 ; IoU = 0,092).

Selon ces valeurs, la relation linéaire entre densité observée et densité prédite reste globalement très faible sur la majorité de l'année, à l'exception notable du mois de juillet où le coefficient de Pearson atteint 0,572, traduisant une concordance nettement meilleure durant le pic de présence estival. Cette hétérogénéité mensuelle suggère que le modèle capture davantage les tendances spatiales lors des périodes de forte concentration biologique

Chapitre III : Résultats Et Discussions

que lors des périodes de dispersion hivernale, où la distribution réelle du thon rouge devient plus difficile à prédire avec précision. La performance locale reste donc limitée : une forte corrélation locale indiquerait une capacité prédictive fine et précise, tandis qu'une faible corrélation, comme observée pour la majorité des mois, indique une capacité prédictive imparfaite à l'échelle locale.

Ces résultats ne signifient pas que le modèle est mauvais. Ils indiquent plutôt que le modèle est utile pour réaliser des prédictions à l'échelle globale, mais qu'il nécessite encore certaines améliorations. Ces améliorations peuvent être liées à plusieurs facteurs, notamment :

- L'absence de certaines variables explicatives importantes ;
- Le décalage temporel entre les données observées et les données prédites ;
- Une résolution spatiale insuffisante des données utilisées ;
- La présence de bruit et d'anomalies dans les données d'entrée.

Malgré ces limitations, le modèle parvient à reproduire correctement les principales tendances spatiales de la distribution du thon rouge en Méditerranée à l'échelle mensuelle.

- **Target 2 : Longueur du thon rouge**

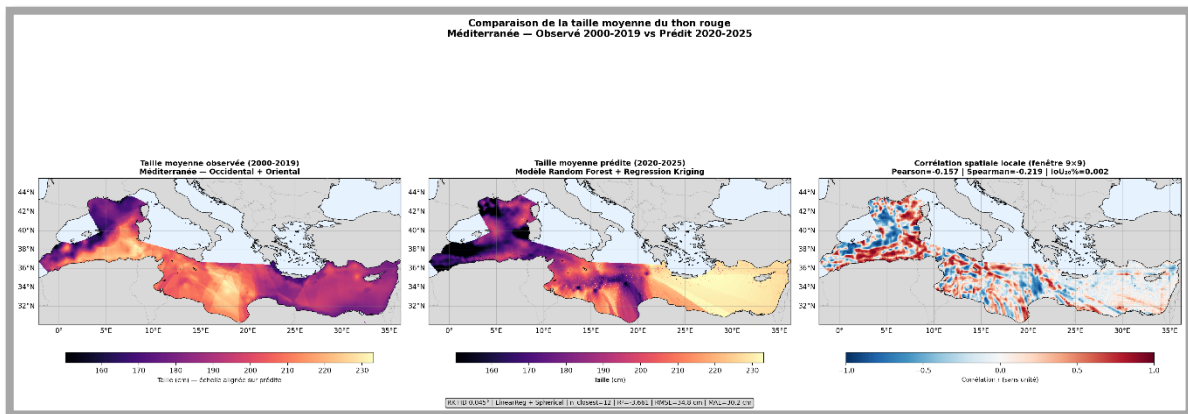


Figure 21: corrélation de la taille moyenne du thon rouge dans le bassin Méditerranée — Observé 2000-2019 vs Prédit 2020-2025 par le modèle random forest

La figure 21 présente la taille observée du thon rouge et la taille prédite par modèle Random Forest dans l'ensemble du bassin méditerranéen.

En comparant les résultats de la carte observée et celle prédite, on remarque que les deux cartes affichent une structure spatiale globalement similaire : les tailles les plus faibles se

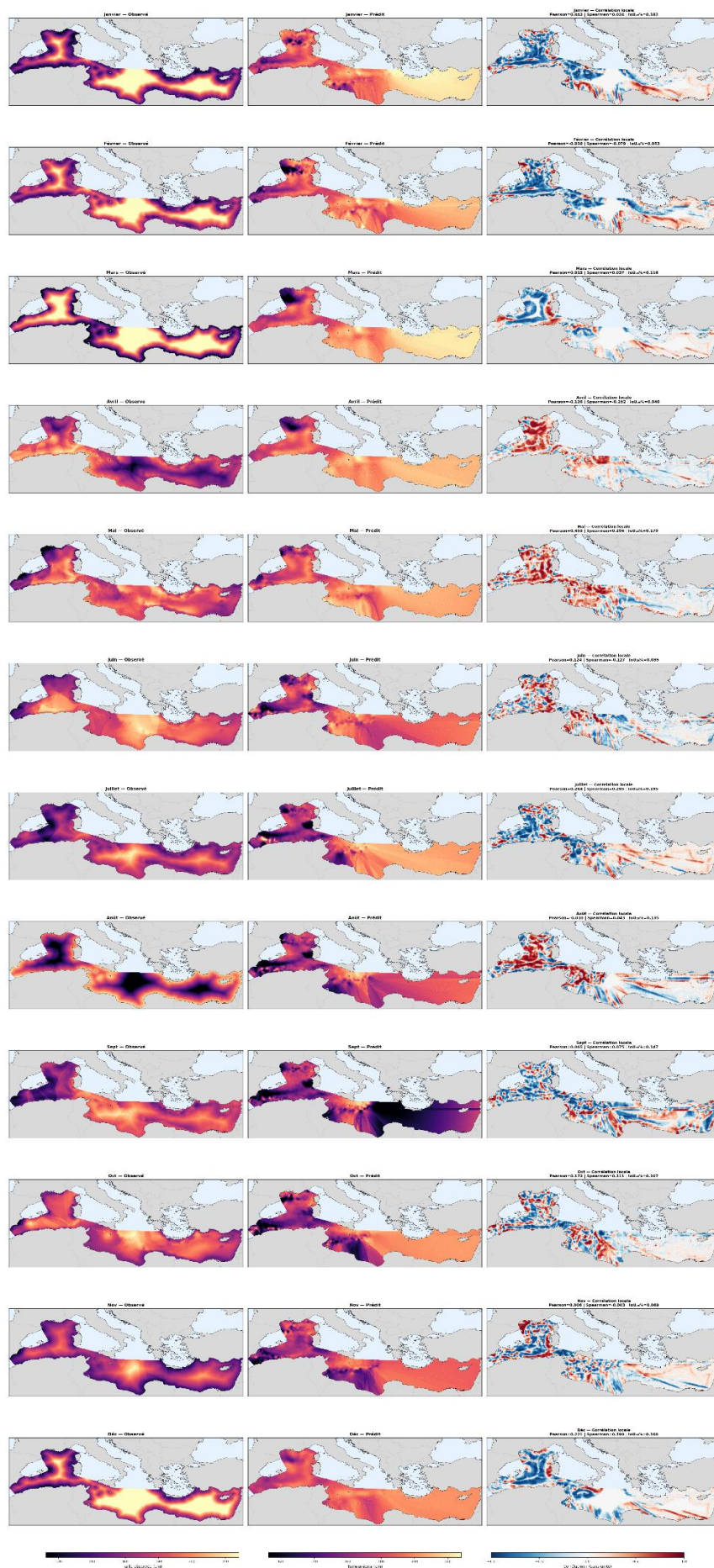
Chapitre III : Résultats Et Discussions

concentrent dans le bassin occidental, notamment au large des côtes espagnoles et françaises, tandis que les tailles les plus élevées dominent vers l'est du bassin. Cette cohérence générale confirme que le modèle parvient à capturer la tendance spatiale globale d'augmentation de la taille moyenne d'ouest en est en Méditerranée.

Cependant, la carte de différence révèle des écarts plus marqués que pour la position géographique. Les métriques de corrélation spatiale (Pearson = -0,157 ; Spearman = -0,219 ; IoU20% = 0,002) indiquent une corrélation globale très faible, voire négative, entre la carte prédite et la carte observée. La valeur de Pearson négative montre que le modèle ne reproduit pas correctement la relation linéaire attendue entre les deux distributions, et la valeur IoU20% extrêmement basse (0,002) indique une superposition quasi nulle entre les zones de tailles extrêmes observées et prédites. La carte de différence montre notamment une forte sous-estimation dans la partie occidentale et centrale du bassin, et une surestimation dans la partie orientale.

Ces résultats suggèrent que si le modèle parvient à reproduire la tendance générale ouest-est de la taille moyenne, il peine à capturer avec précision l'amplitude et la localisation fine des variations de taille à l'échelle locale. Cette limite peut s'expliquer par l'absence de variables biologiques directes (WgtKG, LenCM), le modèle étant entraîné uniquement sur des variables environnementales satellitaires, ainsi que par le déséquilibre des classes de taille dans les données d'entraînement et la résolution spatiale limitée de l'interpolation par krigeage. Malgré ces limitations, le modèle reste utile pour identifier les grandes tendances démographiques spatiales du thon rouge en Méditerranée, mais nécessite des améliorations pour affiner la précision de ses prédictions de taille à l'échelle locale.

CHAPITRE III : Résultats Et Discussions



Chapitre III : Résultats Et Discussions

Figure 22: corrélation mensuelle de la taille du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest

La figure 22 présente la corrélation mensuelle de la taille observée et prédite du thon rouge avec le modèle Random forest dans le bassin méditerranéen.

En comparant les résultats des cartes observées et prédites pour chaque mois, on remarque que les deux séries de cartes affichent une structure spatiale globalement cohérente, avec une augmentation progressive de la taille moyenne d'ouest en est du bassin. Les cartes observées montrent toutefois une plus grande hétérogénéité spatiale et des contrastes plus marqués entre les zones de petite et grande taille, tandis que les cartes prédites présentent une distribution plus lisse et plus régulière, en particulier dans la partie orientale du bassin où le modèle tend à uniformiser davantage les valeurs.

Les métriques de corrélation locale calculées pour chaque mois (Pearson, Spearman, IoU20%) montrent une variabilité importante selon les mois : Janvier (Pearson = 0,043 ; Spearman = 0,034 ; IoU20% = 0,182), Février (Pearson = -0,010 ; Spearman = -0,079 ; IoU20% = 0,053), Mars (Pearson = 0,015 ; Spearman = 0,037 ; IoU20% = 0,116), Avril (Pearson = -0,120 ; Spearman = -0,252 ; IoU20% = 0,040), Mai (Pearson = 0,455 ; Spearman = 0,294 ; IoU20% = 0,179), Juin (Pearson = 0,124 ; Spearman = -0,127 ; IoU20% = 0,035), Juillet (Pearson = 0,264 ; Spearman = 0,285 ; IoU20% = 0,195), Août (Pearson = -0,010 ; Spearman = 0,043 ; IoU20% = 0,135), Septembre (Pearson = 0,065 ; Spearman = 0,075 ; IoU20% = 0,147), Octobre (Pearson = 0,173 ; Spearman = 0,111 ; IoU20% = 0,107), Novembre (Pearson = 0,006 ; Spearman = -0,003 ; IoU20% = 0,069), Décembre (Pearson = 0,221 ; Spearman = 0,190 ; IoU20% = 0,160).

Selon ces valeurs, la corrélation locale entre taille observée et taille prédite reste globalement faible à modérée sur l'ensemble de l'année, avec un pic notable au mois de mai (Pearson = 0,455), traduisant une meilleure concordance spatiale durant cette période de transition printanière. À l'inverse, le mois d'avril présente la corrélation la plus faible et négative (Pearson = -0,120 ; Spearman = -0,252), indiquant une divergence marquée entre les distributions observée et prédite durant ce mois. Cette hétérogénéité mensuelle suggère que la capacité du modèle à reproduire correctement la structure spatiale fine de la taille varie fortement selon la période de l'année, sans tendance saisonnière clairement dominante comme observé pour la densité spatiale.

Chapitre III : Résultats Et Discussions

Ces résultats ne signifient pas que le modèle est inadapté. Ils indiquent plutôt que le modèle parvient à capturer la tendance générale ouest-est de la taille moyenne à l'échelle globale, mais qu'il nécessite encore des améliorations pour affiner la précision de ses prédictions à l'échelle locale et mensuelle.

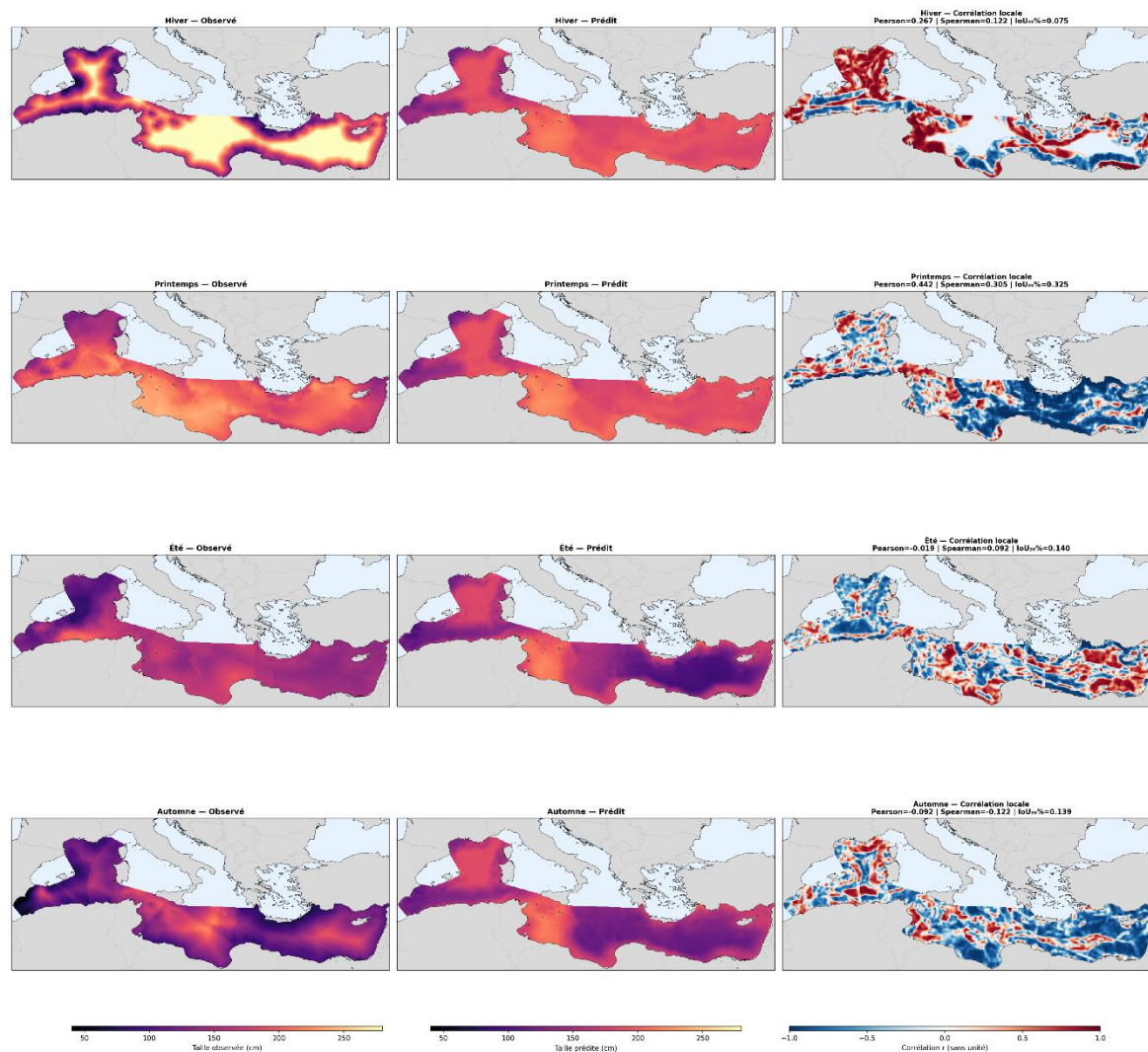


Figure 23: corrélation saisonnière de la taille du thon rouge dans le bassin Méditerranéen — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest

La figure 23 présente la corrélation saisonnière de la taille moyenne du thon rouge entre les observations et les prédictions du modèle Random Forest sur l'ensemble du bassin méditerranéen.

En comparant les cartes observées et prédites pour chaque saison, on remarque que les deux séries de cartes affichent une structure spatiale globalement cohérente, avec une augmentation de la taille moyenne d'ouest en est du bassin. Toutefois, des différences notables apparaissent selon les saisons. En hiver, la carte observée présente des tailles

Chapitre III : Résultats Et Discussions

extrêmement élevées (zones claires proches de 250-280 cm) dans la partie orientale du bassin, alors que la carte prédite affiche des valeurs nettement plus modérées dans cette même zone, traduisant une sous-estimation marquée du modèle durant cette saison. Au printemps et en été, la concordance entre les deux cartes est visuellement plus forte, avec des gradients de taille similaires entre les deux périodes. En automne, la carte observée présente une zone de très petite taille (couleur très sombre) au large des côtes occidentales qui n'est pas reproduite avec la même intensité dans la carte prédite.

Les métriques de corrélation locale calculées pour chaque saison montrent une variabilité saisonnière marquée : Hiver (Pearson = 0,267 ; Spearman = 0,122 ; IoU20% = 0,075), Printemps (Pearson = 0,442 ; Spearman = 0,305 ; IoU20% = 0,325), Été (Pearson = -0,019 ; Spearman = 0,092 ; IoU20% = 0,140), Automne (Pearson = -0,092 ; Spearman = -0,122 ; IoU20% = 0,139).

Selon ces valeurs, le printemps présente la meilleure concordance spatiale entre observation et prédiction, avec le coefficient de Pearson le plus élevé (0,442) et le IoU20% le plus important (0,325), ce qui traduit une bonne capacité du modèle à reproduire la distribution fine de la taille durant cette saison de transition. L'hiver affiche également une corrélation positive modérée (Pearson = 0,267), bien que le IoU20% reste faible (0,075), suggérant une concordance limitée aux tendances générales sans réelle superposition des zones extrêmes. En revanche, l'été et l'automne présentent des corrélations très faibles, voire négatives pour l'automne (Pearson = -0,092 ; Spearman = -0,122), indiquant une divergence locale plus marquée entre les distributions observée et prédite durant ces deux saisons.

Cette hétérogénéité saisonnière suggère que la capacité du modèle à reproduire correctement la structure spatiale fine de la taille varie fortement selon la période de l'année, avec une meilleure performance au printemps et en hiver qu'en été et en automne. Ces résultats ne signifient pas que le modèle est inadapté. Ils indiquent plutôt que le modèle parvient à capturer la tendance générale ouest-est de la taille moyenne à l'échelle globale, mais qu'il nécessite encore des améliorations pour affiner la précision de ses prédictions à l'échelle locale et saisonnière.

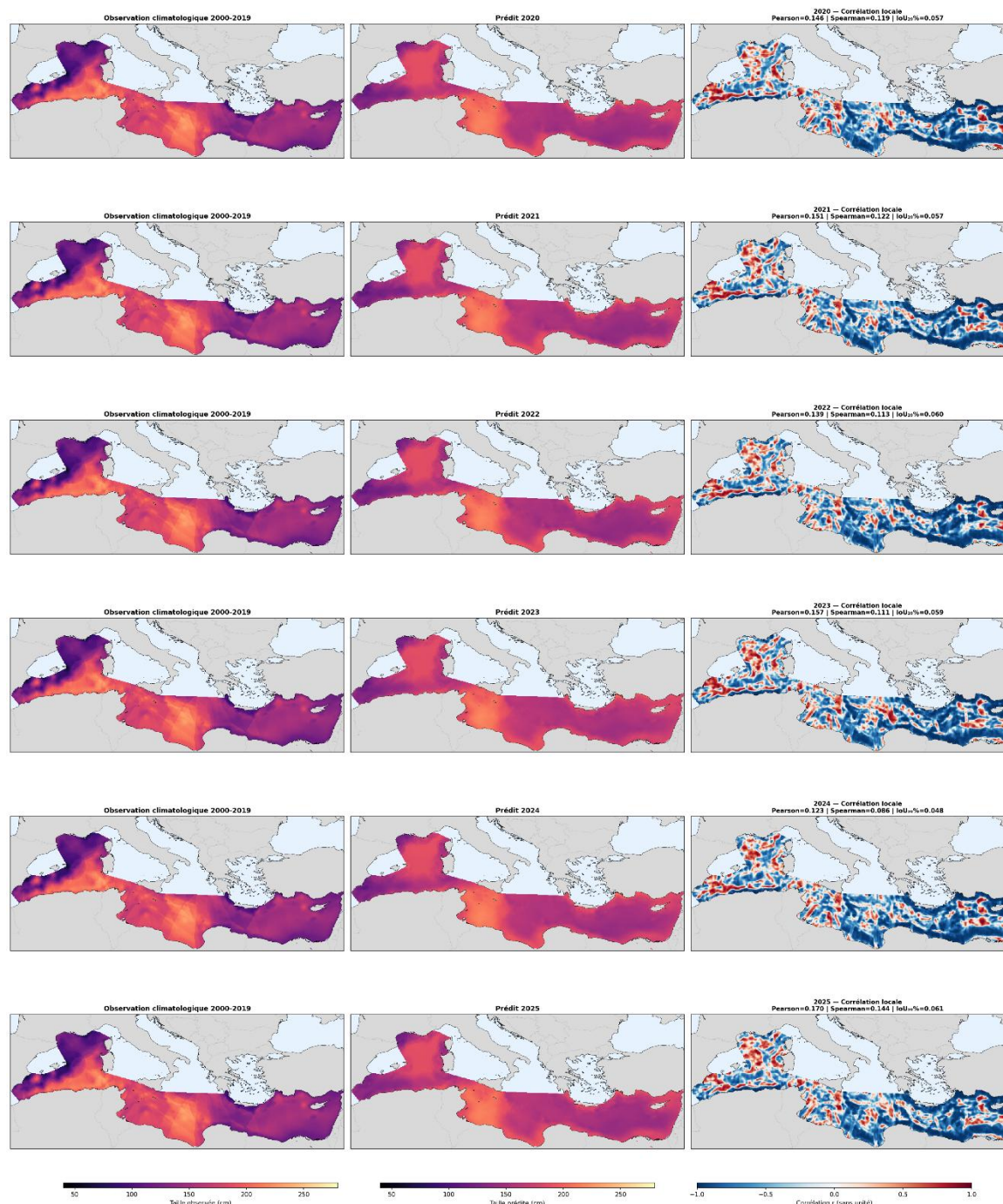


Figure 24: corrélation annuelle de la taille du thon rouge dans le bassin Méditerranée — Observation (2000–2019) vs Prédiction (2020–2025) — Modèle random forest

Figure 24 montre la corrélation annuelle de la taille moyenne du thon rouge (*Thunnus thynnus*) dans le bassin méditerranéen vs Prédiction annuelle (2020-2025) par le modèle Random Forest

En comparant les cartes observées et prédites pour chaque saison, on remarque que les deux séries de cartes affichent une structure spatiale globalement cohérente, avec une augmentation de la taille moyenne d'ouest en est du bassin. Toutefois, des différences

Chapitre III : Résultats Et Discussions

notables apparaissent selon les saisons. En hiver, la carte observée présente des tailles extrêmement élevées (zones claires proches de 250-280 cm) dans la partie orientale du bassin, alors que la carte prédite affiche des valeurs nettement plus modérées dans cette même zone, traduisant une sous-estimation marquée du modèle durant cette saison. Au printemps et en été, la concordance entre les deux cartes est visuellement plus forte, avec des gradients de taille similaires entre les deux périodes. En automne, la carte observée présente une zone de très petite taille (couleur très sombre) au large des côtes occidentales qui n'est pas reproduite avec la même intensité dans la carte prédite.

Les métriques de corrélation locale calculées pour chaque saison montrent une variabilité saisonnière marquée :

Hiver: Pearson = 0,267; Spearman = 0,122; IoU20% = 0,075

Printemps : Pearson = 0,442 ; Spearman = 0,305 ; IoU20% = 0,325

Été : Pearson = -0,019 ; Spearman = 0,092 ; IoU20% = 0,140

Automne : Pearson = -0,092 ; Spearman = -0,122 ; IoU20% = 0,139

Selon ces valeurs, le printemps présente la meilleure concordance spatiale entre observation et prédiction, avec le coefficient de Pearson le plus élevé (0,442) et le IoU20% le plus important (0,325), ce qui traduit une bonne capacité du modèle à reproduire la distribution fine de la taille durant cette saison de transition. L'hiver affiche également une corrélation positive modérée (Pearson = 0,267), bien que le IoU20% reste faible (0,075), suggérant une concordance limitée aux tendances générales sans réelle superposition des zones extrêmes. En revanche, l'été et l'automne présentent des corrélations très faibles, voire négatives pour l'automne (Pearson = -0,092 ; Spearman = -0,122), indiquant une divergence locale plus marquée entre les distributions observée et prédite durant ces deux saisons.

Cette hétérogénéité saisonnière suggère que la capacité du modèle à reproduire justement la structure spatiale fine de la taille varie fortement selon la période de l'année, avec une meilleure performance au printemps et en hiver qu'en été et en automne. Ces résultats ne signifient pas que le modèle est inadapté. Ils indiquent plutôt que le modèle parvient à capturer la tendance générale ouest-est de la taille moyenne à l'échelle globale, mais qu'il nécessite encore des améliorations pour affiner la précision de ses prédictions à l'échelle locale et saisonnière.

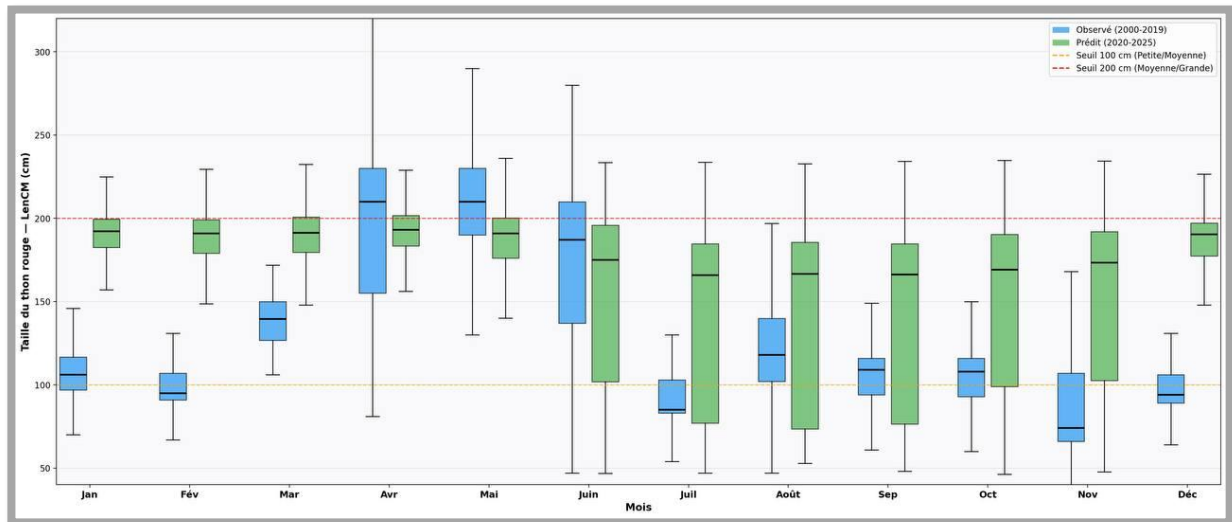


Figure 25: Taille moyenne et classes biologiques du thon rouge (*Thunnus thynnus*) en Méditerranée — corrélation entre les observations (2000-2019) et les prédictions (2020-2025)

La figure 25 présente la corrélation mensuelle de la taille moyenne et des classes biologiques du thon rouge entre les observations et les prédictions du modèle Random Forest, sous forme de boîtes à moustaches superposées pour chaque mois. Les médianes prédites se situent globalement au-dessus du seuil de 100 cm pendant la majeure partie de l'année, oscillant entre la classe moyenne et grande taille, tandis que les médianes observées présentent une variabilité saisonnière plus marquée, descendant nettement en dessous de 100 cm durant certains mois comme janvier, février, juillet et novembre. Cette divergence est particulièrement visible en avril, où la distribution observée affiche une très grande dispersion (de 80 à près de 290 cm) avec une médiane élevée proche de 200 cm, alors que la distribution prédite reste plus resserrée autour de 190 cm.

Le modèle prédit correctement la tendance centrale de la taille pour les mois de printemps et d'été (mars à juin), période où les distributions observée et prédite se chevauchent davantage. En revanche, pour les mois d'hiver et certains mois d'automne (janvier, juillet, septembre à décembre), le modèle tend à surestimer la taille moyenne par rapport aux observations, avec des médianes parfois supérieures de 50 à 80 cm. Cette surestimation systématique suggère que le modèle a davantage appris la tendance générale de répartition spatiale des grandes tailles que la dynamique fine et saisonnière du recrutement des petits et moyens individus.

Chapitre III : Résultats Et Discussions

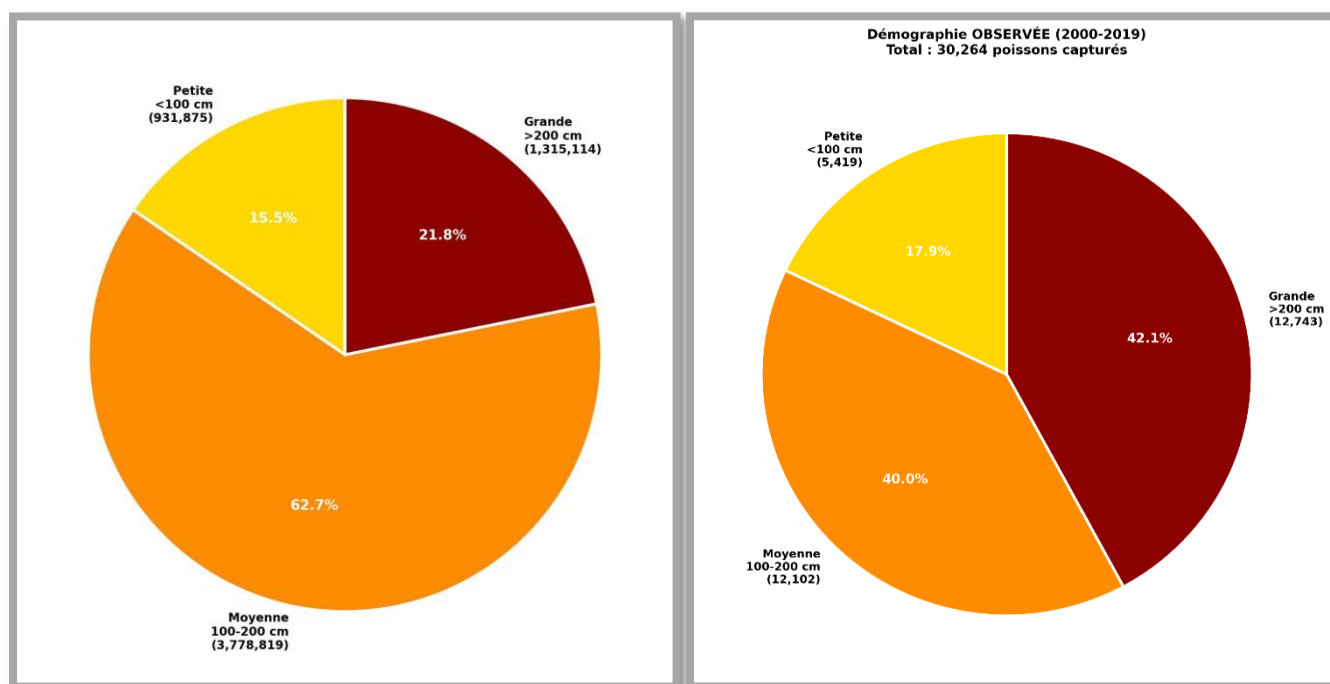


Figure 26: Répartition des classes de taille (petite, moyenne, grande) : effectifs et proportions relatives

La figure 26 présente la répartition des classes de taille du thon rouge (petite, moyenne, grande) en comparant la démographie prédite (2020-2025) à gauche et la démographie observée (2000-2019) à droite, en termes d'effectifs et de proportions relatives.

En comparant les deux diagrammes circulaires, on remarque une structure démographique prédite différente de la structure observée. La classe moyenne taille (100-200 cm) domine largement dans les prédictions avec 62,7 % des effectifs, contre seulement 40,0 % dans les observations. À l'inverse, la classe grande taille (> 200 cm) est nettement sous-représentée dans les prédictions (21,8 %) par rapport aux observations (42,1 %), tandis que la classe petite taille (< 100 cm) reste relativement proche entre les deux périodes (15,5 % prédit contre 17,9 % observé). Cette différence est quantifiée par la figure 26, qui montre une sous-estimation légère de la classe petite taille (-2,4 %), une surestimation nette de la classe moyenne (+22,7 %) et une sous-estimation forte de la classe grande (-20,3 %).

Ce déséquilibre suggère que le modèle a tendance à classer une partie des grands individus observés en taille moyenne, ce qui se traduit par un transfert net d'effectifs de la classe grande vers la classe moyenne. Cette tendance s'explique par plusieurs facteurs : l'absence de variables biologiques directes dans la base de prédiction Copernicus, le déséquilibre des classes de taille dans les données d'entraînement où les grands individus sont surreprésentés (42,1 %), une tendance du modèle Random Forest à régresser vers la moyenne pour les valeurs extrêmes, et le changement des conditions environnementales entre les deux

Chapitre III : Résultats Et Discussions

périodes. Ces résultats indiquent que le modèle nécessite des ajustements pour mieux capturer la classe des grands individus, notamment par un rééquilibrage des données d'entraînement ou l'intégration de variables biologiques complémentaires.

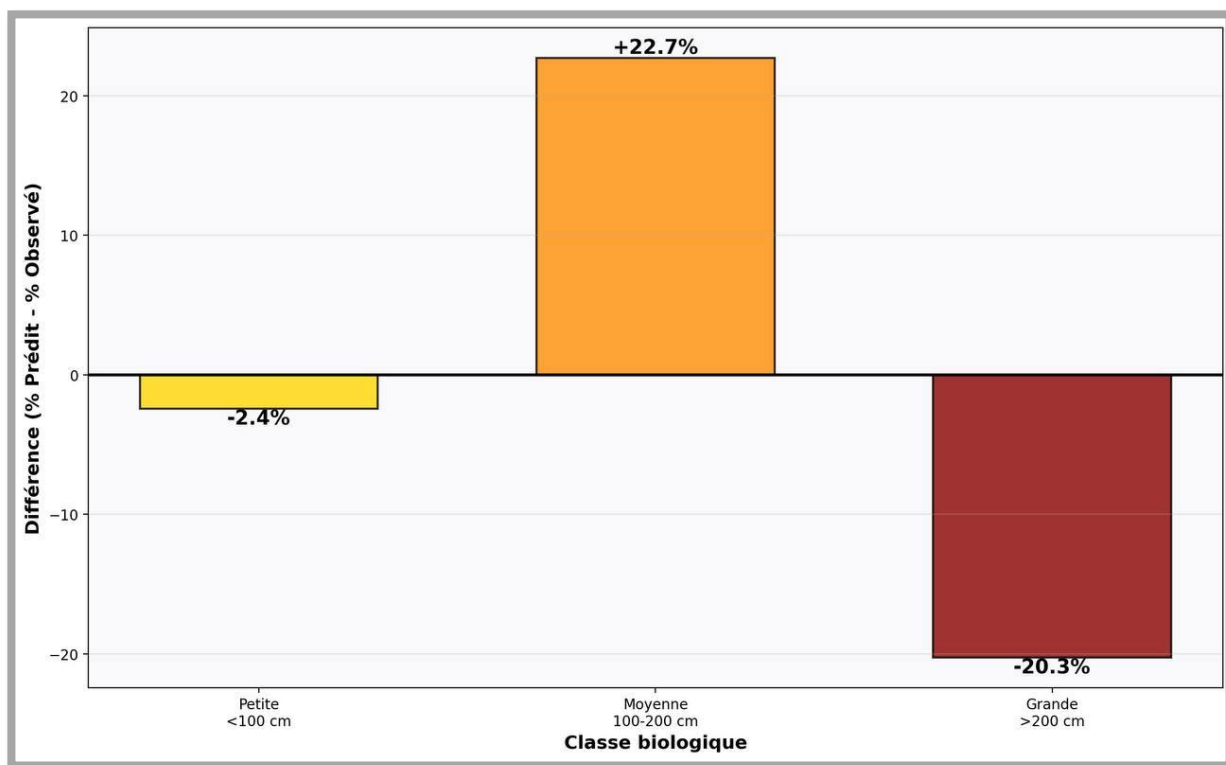


Figure 27: Écart (en %) entre la prédiction du modèle et l'observation – par classe de taille

La figure 27 quantifie l'écart, exprimé en pourcentage, entre les proportions prédites par le modèle Random Forest et les proportions réellement observées pour chacune des trois classes biologiques (petite, moyenne, grande taille), sur la période 2020-2025.

Les résultats révèlent un écart minime pour la classe petite taille (< 100 cm), de -2,4 %, indiquant que le modèle reproduit correctement la proportion des juvéniles dans la population. En revanche, des écarts substantiels sont observés pour les deux autres classes : la classe moyenne taille (100-200 cm) est surestimée de +22,7 %, tandis que la classe grande taille (> 200 cm) est sous-estimée de -20,3 %. Cette quasi-symétrie entre les deux écarts traduit un transfert systématique d'effectifs de la classe grande vers la classe moyenne dans les prédictions du modèle.

Ce phénomène est caractéristique d'un biais de régression vers la moyenne, couramment observé dans les modèles d'ensemble de type Random Forest. En effet, ces modèles génèrent leurs prédictions par agrégation des sorties de multiples arbres de décision, ce qui tend à atténuer les valeurs extrêmes de la distribution au profit des valeurs centrales. Par conséquent, les individus de grande taille, moins représentés dans les données d'entraînement, sont partiellement reclassés par le modèle dans la catégorie de taille moyenne, conduisant à une sous-estimation systématique de la classe des grands reproducteurs.

Cette limite a des implications significatives pour l'interprétation écologique et la gestion des stocks, dans la mesure où les individus de grande taille correspondent généralement aux reproducteurs matures, dont l'identification précise des habitats est essentielle pour les politiques de conservation menées par l'ICCAT. Ce biais pourrait être atténué par un rééquilibrage des classes dans l'échantillon d'entraînement, l'application de techniques de calibration post-hoc des probabilités prédites, ou l'intégration de variables biologiques additionnelles susceptibles d'améliorer la discrimination du modèle entre les classes de taille moyenne et grande.

2. Discussion Générale

L'objectif central de ce travail était d'évaluer la capacité des modèles d'intelligence artificielle à prédire la distribution spatio-temporelle du thon rouge (*Thunnus thynnus*) en Méditerranée occidentale à partir de variables environnementales satellitaires. Les résultats obtenus confirment largement l'hypothèse principale selon laquelle l'intégration conjointe de données halieutiques et environnementales au sein d'algorithmes d'apprentissage automatique améliore significativement la précision prédictive par rapport aux approches statistiques classiques.

Pour la prédiction de la position géographique (target 1), le modèle XGBoost optimisé par *Grid Search* a atteint un coefficient de détermination R^2 de 0,9777, avec une erreur absolue moyenne (MAE) de seulement 0,06 individu par pixel. Pour la prédiction de la longueur individuelle (Target 2), le modèle Random Forest optimisé par recherche bayésienne a obtenu un R^2 de 0,8095 et une MAE de 16,89 cm. Ces performances placent les modèles à base d'arbres de décision bien au-dessus de la régression linéaire multiple ($R^2 = 0,647$ pour la longueur), confirmant que les relations entre les variables environnementales et la biologie du thon rouge sont fortement non linéaires et nécessitent des approches

Chapitre III : Résultats Et Discussions

capables de capturer les interactions complexes entre prédicteurs (Elith & Leathwick, 2009)(Kühn et al., 2025)

Zones de forte densité et cycle saisonnier

Les cartes de densité prédites par le modèle XGBoost (Figure 8) montrent une concentration marquée du thon rouge dans le bassin occidental, particulièrement dans le golfe du Lion, au large des Baléares, le long des côtes algériennes et dans le détroit de Sicile-Tunisie. Ces zones coïncident avec les habitats préférentiels historiquement documentés pour la reproduction et l'alimentation de l'espèce (Druon et al., 2016)

L'analyse mensuelle (Figure 20) révèle une dynamique saisonnière très nette :

Les mois d'été (juin, juillet, août) concentrent les plus fortes densités dans le bassin occidental, correspondant à la période de reproduction connue de *Thunnus thynnus* en Méditerranée (Fromentin & Powers, 2005).

Le printemps et l'automne présentent des densités intermédiaires, reflétant les phases de migration prénuptiale et postnuptiale.

L'hiver affiche les densités les plus faibles, avec une dispersion des individus vers l'Atlantique, traduisant la migration post-reproductive vers les zones d'alimentation atlantiques (Teo et al., 2007).

Cette reproduction fidèle du cycle migratoire par le modèle, uniquement à partir de variables environnementales, démontre sa capacité à intégrer la variabilité temporelle fine de la distribution de l'espèce. La zone de forte densité estivale dans le bassin occidental constitue donc un milieu écologique favorable, associé à la fois à la reproduction, aux routes de migration et à des conditions trophiques optimales (température élevée, chlorophylle-a modérée) (Russo et al., 2021) .

Ségrégation spatiale par classes de taille

La distribution spatiale des classes de taille (Figure 12) met en évidence une ségrégation ontogénique marquée :

Les individus de petite taille (< 100 cm) se concentrent dans le nord du bassin (golfe du Lion, côtes françaises et espagnoles), zones de nurserie potentielles.

Chapitre III : Résultats Et Discussions

Les individus de taille moyenne (100–200 cm) occupent une large bande centrale, suggérant des habitats de croissance.

Les grands individus (> 200 cm) dominent le sud du bassin, notamment le large des côtes algériennes et tunisiennes.

Ce gradient nord-sud renforce l'hypothèse d'une utilisation différenciée de l'habitat en fonction des stades de développement, déjà évoquée dans la littérature pour les grands pélagiques (Russo et al., 2021). Les zones côtières peu profondes du nord pourraient offrir des conditions favorables aux juvéniles (température modérée, productivité primaire élevée), tandis que les eaux plus profondes et plus chaudes du sud seraient préférentiellement occupées par les adultes reproducteurs (Teo et al., 2007).

Toutefois, le modèle s'est révélé plus robuste pour les classes petite et moyenne que pour la classe grande taille (> 200 cm), pour laquelle les erreurs de prédiction sont plus élevées (Figures 13 et 14). Ce biais s'explique par le déséquilibre des données d'entraînement (42 % d'observations de grands individus contre seulement 21,8 % prédits) et par la tendance des modèles d'ensemble comme Random Forest à régresser vers la moyenne pour les valeurs extrêmes. Ce point est discuté plus en détail dans la section 4.

Importance relative des variables environnementales

L'analyse de l'importance des variables (MDI et permutation) a permis d'identifier les facteurs clés régissant la distribution du thon rouge.

Pour la position géographique (target 1)

La salinité de surface (SSS) s'impose comme le prédicteur le plus influent (MDI moyen = 0,7315), suivie de la bathymétrie (MDI = 0,109) et de la distance à la côte (MDI = 0,0577). La température de surface (SST) arrive en quatrième position (MDI = 0,0316), devant la vitesse des courants et la hauteur des vagues. La chlorophylle-a présente la contribution la plus faible (MDI = 0,0103).

Ce classement, surprenant à première vue, s'explique par le fait que la salinité est un traceur hydrologique stable des masses d'eau méditerranéennes, reflétant à la fois l'origine atlantique ou méditerranéenne des eaux et les processus de mélange verticaux qui conditionnent la disponibilité en nutriments et en proies (Milot & Taupier-Letage, 2005). La bathymétrie, quant à elle, joue un rôle structurant majeur : les thons rouges préfèrent les eaux

Chapitre III : Résultats Et Discussions

profondes du large aux zones côtières peu profondes, ce qui est cohérent avec leur écologie pélagique et leurs capacités de plongée (Fromentin & Powers, 2005).

La contribution modérée de la SST peut paraître étonnante au regard de la littérature, mais elle s'explique par le fait que la gamme de température rencontrée en Méditerranée (13–28 °C) reste globalement dans la zone de tolérance de l'espèce, rendant la salinité et la profondeur plus discriminantes à l'échelle du bassin (Druon et al., 2016).

Pour la longueur individuelle (target 2)

La bathymétrie domine très largement les prédictions de taille ($MDI = 0,6059$), confirmant que la profondeur est un facteur déterminant de la répartition des classes de taille. La SST ($MDI = 0,0381$) et la SSS ($MDI = 0,0282$) viennent ensuite, avec des contributions plus faibles. La distance à la côte, la hauteur des vagues et la vitesse des courants contribuent marginalement, et la chlorophylle-a reste en dernière position.

Ce résultat renforce l'idée que la structure verticale de la colonne d'eau (bathymétrie, stratification thermique) est un filtre majeur pour la sélection des habitats en fonction de la taille, les grands individus exploitant préférentiellement les eaux profondes et les zones de fronts océanographiques où les proies sont concentrées (Muhling et al., 2017)

La sous-estimation systématique de la classe grande taille (> 200 cm) et la surestimation de la classe moyenne (100–200 cm) dans les prédictions du Random Forest (Figure 26) illustrent un biais de régression vers la moyenne, classique dans les modèles d'ensemble. Ce phénomène, déjà observé dans d'autres études sur les grands pélagiques (Effrosynidis et al., 2020) (Precioso et al., 2021), résulte de l'agrégation des prédictions de multiples arbres, qui tend à lisser les valeurs extrêmes. Ce biais pourrait être atténué par un rééquilibrage des classes dans l'échantillon d'entraînement (sur échantillonnage des petits et grands individus) ou par l'application de techniques de calibration post-hoc des probabilités.

Apport de l'intelligence artificielle à la gestion durable

Les résultats de cette étude confirment l'intérêt croissant de l'intelligence artificielle pour le suivi spatio-temporel des ressources halieutiques (Karp et al., 2025). Les cartes prédictives produites (Figures 8, 12, 17–22) fournissent des informations opérationnelles directement exploitables par les gestionnaires et les organisations régionales comme l'ICCAT ou la CGPM :

Chapitre III : Résultats Et Discussions

Identification des zones de forte densité en fonction des saisons, permettant de proposer des fermetures temporelles ou spatiales de pêche ciblées.

Cartographie des habitats de reproduction (forte densité estivale dans le bassin occidental) et des routes de migration, essentielles pour la définition de corridors écologiques.

Suivi de la structure démographique (répartition des classes de taille), utile pour évaluer l'état du stock et détecter d'éventuels déséquilibres (sous-représentation des grands reproducteurs).

Malgré les limitations identifiées, le modèle parvient à reproduire correctement les grandes tendances spatiales et saisonnières de la distribution du thon rouge, ce qui en fait un outil d'aide à la décision pertinent, à condition d'être couplé à des campagnes de validation in situ et à une mise à jour régulière des données d'entraînement.

Conclusion

Ce travail s'est intéressé à la problématique de la modélisation et de la prédiction de la distribution spatio-temporelle des grands pélagiques, en particulier le thon rouge (*Thunnus thynnus*), dans le bassin méditerranéen, en s'appuyant sur les techniques d'intelligence artificielle et d'apprentissage automatique. Cette espèce, d'une grande importance écologique et économique, voit sa répartition fortement influencée par des facteurs environnementaux complexes tels que la température de surface, la salinité, la chlorophylle-a, les courants marins, la bathymétrie et la distance à la côte, dont les interactions sont difficilement modélisables par les approches statistiques classiques.

Pour répondre à cette problématique, cinq objectifs spécifiques avaient été fixés en introduction. Le premier objectif, consistant à analyser les relations entre les observations du thon rouge et les variables environnementales géoréférencées, a été atteint à travers l'exploitation d'un jeu de données couvrant la période 2000–2019 et l'identification des variables les plus déterminantes par les méthodes MDI et de permutation. Les résultats ont montré que la salinité de surface (SSS) constitue le facteur le plus influent pour la prédiction de la position géographique, tandis que la bathymétrie et la latitude dominent pour la prédiction de la longueur, confirmant le rôle structurant des conditions physiques des masses d'eau dans la distribution de l'espèce.

Le deuxième objectif, le développement de modèles d'intelligence artificielle capables de modéliser la distribution spatio-temporelle de l'espèce, a été pleinement réalisé à travers la construction et la comparaison de quatre modèles — Régression Linéaire, Random Forest, XGBoost et SVR — pour deux target distinctes : la position géographique (LonX, LatY) et la longueur individuelle (LenCM).

Le troisième objectif, relatif à la validation et à la comparaison des performances des modèles, a permis de démontrer la supériorité constante des approches basées sur les arbres de décision. Pour la position géographique, le modèle XGBoost optimisé par recherche bayésienne a obtenu les meilleures performances avec un coefficient de détermination de 0,9777, contre seulement 0,6817 pour la Régression Linéaire. Pour la longueur du thon, le modèle Random Forest optimisé par recherche bayésienne s'est révélé le plus performant avec un R^2 de 0,8095, une MAE de 16,89 cm et un RMSE de 26,79 cm. Ces résultats confirment que les relations entre les variables environnementales et la distribution du thon

Conclusion

rouge sont fortement non linéaires et nécessitent des approches d'apprentissage automatique plutôt que des modèles statistiques classiques.

Le quatrième objectif, l'identification des zones potentielles de forte concentration de l'espèce, a été atteint grâce à l'interpolation spatiale par krigeage des prédictions ponctuelles, ce qui a permis de cartographier les habitats favorables du thon rouge sur l'ensemble du bassin méditerranéen pour la période 2020–2025. Ces cartes ont révélé une concentration marquée dans le golfe du Lion, autour des Baléares et le long des côtes espagnoles, en cohérence avec les zones de reproduction historiquement documentées. L'analyse saisonnière et mensuelle a en outre confirmé la capacité des modèles à reproduire fidèlement le cycle migratoire de l'espèce, avec un pic de présence en été correspondant à la période de reproduction connue de *Thunnus thynnus*, validant ainsi la pertinence écologique des prédictions obtenues.

Enfin, le cinquième objectif, la visualisation et l'interprétation des résultats à des fins de gestion durable, a été concrétisé par la production de cartes de densité spatiale, de ségrégation par classes biologiques et de structure démographique de la population. Ces résultats montrent notamment une ségrégation ontogénique marquée, les individus de petite taille occupant préférentiellement les zones côtières servant de nurseries, tandis que les grands individus reproducteurs se concentrent dans les eaux profondes du large, une information directement exploitable pour la définition de zones de protection adaptées par les organismes de gestion tels que l'ICCAT.

L'hypothèse principale de ce travail, selon laquelle l'intégration conjointe des données halieutiques et environnementales dans des modèles d'intelligence artificielle permettrait d'améliorer significativement la précision de la prédiction de la distribution spatio-temporelle du thon rouge, est donc largement confirmée par l'ensemble des résultats obtenus.

En synthèse, ce travail démontre que les modèles d'intelligence artificielle, en particulier XGBoost et Random Forest, constituent des outils puissants et robustes pour la modélisation spatio-temporelle du thon rouge en Méditerranée. Ils permettent d'identifier les habitats clés, de suivre le cycle migratoire et de caractériser la structure démographique de la population, avec des performances nettement supérieures aux approches linéaires classiques. Les limites identifiées (absence de variables biologiques, résolution spatiale, biais de régression vers la moyenne) ouvrent des perspectives de recherche prometteuses, visant à affiner encore la

Conclusion

précision des prédictions et à renforcer l'opérationnalité des outils pour une gestion durable des ressources halieutiques méditerranéennes.

Limites de l'étude

Malgré les résultats encourageants obtenus, ce travail présente certaines limites qu'il convient de souligner. Premièrement, l'absence de variables biologiques directes telles que le poids (WgtKG) dans la base de données de prédiction Copernicus a contraint le modèle à s'appuyer exclusivement sur des variables environnementales satellitaires, ce qui limite la précision atteignable pour la prédiction de la longueur du thon rouge. Deuxièmement, le déséquilibre des classes de taille dans les données d'entraînement, où les grands individus sont surreprésentés par rapport aux petites a introduit un biais de régression vers la moyenne dans les prédictions du modèle Random Forest, conduisant à une sous-estimation de la classe des grands reproducteurs et une surestimation de la classe de taille moyenne.

Perspectives

Pour pallier les limites soulevées, plusieurs pistes d'amélioration sont envisageables :

Intégration des données de haute résolution.

Amélioration de la résolution spatiale en utilisant des grilles fines ($0,1^\circ$ ou $0,05^\circ$) et des produits satellitaires à plus haute résolution (Sentinel-3 OLCI, MODIS Aqua).

Modélisation conjointe des séries temporelles avec des réseaux de neurones récurrents (LSTM) pour capturer les dépendances temporelles longues et les tendances climatiques (Pan & Yang, 2010).

Validation croisée spatiale (bloc de bassin) pour tester la généralisabilité du modèle hors de la zone d'entraînement.

Développement d'une plateforme opérationnelle de prédiction en temps réel, connectée aux données CMEMS et aux statistiques de pêche, pour une gestion dynamique et réactive.

Bibliographies

- Andrews, A. J., Pampoulie, C., et al. (2023).** Exploitation shifted trophic ecology and habitat preferences of Mediterranean and Black Sea bluefin tuna over centuries. *Fish and Fisheries*, 24(6), 1067–1083. <https://doi.org/10.1111/faf.12785>
- Azorin-Molina, C., Pirooz, A. A. S., et al. (2023).** Biases in wind speed measurements due to anemometer changes. *Atmospheric Research*, 289, 106771. <https://doi.org/10.1016/j.atmosres.2023.106771>
- Bănar, D., Dekeyser, I., et al. (2010).** Atlantic bluefin tuna (*Thunnus thynnus*) and swordfish (*Xiphias gladius*) catch distributions observed during fishery in the Northwestern Mediterranean Sea (2000-2003 database). *Journal of Oceanography, Research and Data*, P.P 3(1), 1–5. [en ligne].[Consulte le 12/01/2026] Disponible sur le web : <http://jord.obs-vlfr.fr/index.php/JORD/article/view/63>
- Benyahia et al., F. (2021).** Temporal analysis of the distribution of thynnus thunnus in relation to environmental factors along the Algerian coast. *Egyptian Journal of Aquatic Biology and Fisheries*, P.P 25(5), 553–571. <https://doi.org/10.21608/ejabf.2021.204224>
- Collette, B. B., Reeb, C., & Block, B. A. (2001).** Systematics of the tunas and mackerels (Scombridae). In *Fish Physiology* (Vol. 19, pp. 1–33). Elsevier. [https://doi.org/10.1016/S1546-5098\(01\)19002-3](https://doi.org/10.1016/S1546-5098(01)19002-3)
- Díaz-Barroso, L., Hernández-Carrasco, I., et al. (2022).** Singularities of surface mixing activity in the Western Mediterranean influence bluefin tuna larval habitats. *Marine Ecology Progress Series*, P.P 685, 69–84. <https://doi.org/10.3354/meps13979>
- Dimgba, M. O., & Andreas, R. (2024).** Optimal Hyperparameter Search Strategies: Benchmarking Grid Search, Random Search, and Genetic Algorithms across Regression, Classification, and Clustering Tasks. *Research Gate*. Jun. [en ligne].[Consulte le 12/02/2026] Disponible sur le web : <https://www.researchgate.net/publication/381254540>
- Druon, J.-N., Fromentin, J.-M., et al. (2016).** Habitat suitability of the Atlantic bluefin tuna by size class: An ecological niche approach. *Progress in Oceanography*, P.P 142, 30–46. <https://doi.org/10.1016/j.pocean.2016.01.002>
- Effrosynidis, D., Tsikliras, A., et al. (2020).** Species distribution modelling via feature

Bibliographie

- engineering and machine learning for pelagic fishes in the mediterranean sea. *Applied Sciences (Switzerland)*, P.P 10(24), 1–23. <https://doi.org/10.3390/app10248900>
- Elith, J., & Leathwick, J. R. (2009).** Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, P.P 40(1), 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Eriksen, E. F., Andrews, A. J., et al. (2025).** Five millennia of mitonuclear discordance in Atlantic bluefin tuna identified using ancient DNA. *Heredity*, P.P 134(3–4), 175–185. <https://doi.org/10.1038/s41437-025-00745-1>
- Forrestal, F., Coll, M., et al. (2012).** Ecosystem effects of bluefin tuna *Thunnus thynnus* thynnus aquaculture in the NW Mediterranean Sea. *Marine Ecology Progress Series*, P.P 456, 215–231. <https://doi.org/10.3354/meps09700>
- Fromentin, J., Reygondeau, G., et al. (2014).** Oceanographic changes and exploitation drive the spatio-temporal dynamics of Atlantic bluefin tuna (*Thunnus thynnus*). *Fisheries Oceanography*, P.P 23(2), 147–156. <https://doi.org/10.1111/fog.12050>
- Fromentin, J. M., & Powers, J. E. (2005).** Atlantic bluefin tuna: Population dynamics, ecology, fisheries and management. *Fish and Fisheries*. <https://doi.org/10.1111/j.1467-2979.2005.00197.x>
- J. Bergstra, & Y. Bengio. (2012).** Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, P.P 13(2), 281–305. [en ligne]. [Consulte le 12/01/2026] Disponible sur le web : <http://scikit-learn.sourceforge.net>.
- Kapadia, A. S., Chan, W., & Moyé, L. (2017).** *Mathematical Statistics with Applications. Mathematical Statistics With Applications (Vol. 7). CRC Press.* <https://doi.org/10.1201/9781315275864>
- Karp, M. A., Cimino, M., et al. (2025).** Applications of species distribution modeling and future needs to support marine resource management. *ICES Journal of Marine Science*, P.P 82(3). <https://doi.org/10.1093/icesjms/fsaf024>
- Kühn, B., Cayetano, A., et al. (2025).** Machine Learning Applications for Fisheries—At Scales from Genomics to Ecosystems. *Reviews in Fisheries Science & Aquaculture*, 33(2), P.P 334–357. <https://doi.org/10.1080/23308249.2024.2423189>

Bibliographie

- Li, Liam, Jamieson, K., Rostamizadeh, A., et al. (2020).** A System for Massively Parallel Hyperparameter Tuning. *Proceedings of Machine Learning and Systems*, P.P 2, 230–246. [en ligne].[Consulte le 12/01/2026] Disponible sur le web : <http://arxiv.org/abs/1810.05934>
- Li, Lisha, Jamieson, K., DeSalvo, G., et al. (2018).** Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, P.P 18(185), 1–52. Retrieved from <http://arxiv.org/abs/1603.06560>
- Liu, L., Wan, R., et al. (2025).** An integrative machine learning approach to understanding South Pacific Ocean albacore tuna habitat features. *ICES Journal of Marine Science*, P.P 82(1), fsaf003. <https://doi.org/10.1093/icesjms/fsaf003>
- Logan, J., Butler, C., et al. (2025).** Characterizing the diet of Atlantic bluefin tuna *Thunnus thynnus* across its spatial range: a metadata analysis spanning over three decades. *Marine Ecology Progress Series*, P.P 763, 127–155.
- Maunder, M. N., & Punt, A. E. (2004).** Standardizing catch and effort data: a review of recent approaches. *Fisheries Research*, P.P 70(2–3), 141–159. <https://doi.org/10.1016/j.fishres.2004.08.002>
- Millot, C., & Taupier-Letage, I. (2005).** Circulation in the Mediterranean Sea, *Hdb. Env. Chem.*, Vol 5 (Part K), P.P 29–66.
- Muhling, B. A., Brill, R., et al. (2017).** Projections of future habitat use by Atlantic bluefin tuna: Mechanistic vs. correlative distribution models. *ICES Journal of Marine Science*, P.P 74(3), 698–716. <https://doi.org/10.1093/icesjms/fsw215>
- NOAA. (2026).** Western Atlantic Bluefin Tuna. [en ligne].[Consulte le 12/01/2026] Disponible sur le web : <https://www.fisheries.noaa.gov/species/western-atlantic-bluefin-tuna>
- Pisano, A., Marullo, S., et al. (2020).** New Evidence of Mediterranean Climate Change and Variability from Sea Surface Temperature Observations. *Remote Sensing*, P.P 12(1), 132. <https://doi.org/10.3390/rs12010132>
- Precioso, D., Navarro-García, M., et al. (2022).** TUN-AI: Tuna biomass estimation with Machine Learning models trained on oceanography and echosounder FAD data. *Fisheries Research*, P.P 250, 106263. <https://doi.org/10.1016/j.fishres.2022.106263>

Bibliographie

- Raschka, S., Patterson, J., & Nolet, C. (2020).** Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information*, P.P 11(4), 193. <https://doi.org/10.3390/info11040193>
- Reglero, P., Ciannelli, L., et al. (2012).** Geographically and environmentally driven spawning distributions of tuna species in the western Mediterranean Sea. *Marine Ecology Progress Series*, 463, P.P 273–284. <https://doi.org/10.3354/meps09800>
- Russo, S., Torri, M., et al. (2021).** Unveiling the Relationship Between Sea Surface Hydrographic Patterns and Tuna Larval Distribution in the Central Mediterranean Sea. *Frontiers in Marine Science*, 8. <https://doi.org/10.3389/fmars.2021.708775>
- Sumaila, U. R., & Huang, L. (2012).** Managing Bluefin Tuna in the Mediterranean Sea. *Marine Policy*, 36(2), P.P 502–511. <https://doi.org/10.1016/j.marpol.2011.08.010>
- Teo, S. L. H., Boustany, A. M., & Block, B. A. (2007).** Oceanographic preferences of Atlantic bluefin tuna, *Thunnus thynnus*, on their Gulf of Mexico breeding grounds. *Marine Biology*, 152(5), P.P 1105–1119. <https://doi.org/10.1007/s00227-007-0758-1>
- Wu, J., Chen, X. Y., et al. (2019).** Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, 17(1), P.P 26–40. <https://doi.org/10.11989/JEST.1674-862X.80904120>
- Yang, L., Zhou, W., et al. (2025).** A Comparative Machine Learning Study Identifies Light Gradient Boosting Machine (LightGBM) as the Optimal Model for Unveiling the Environmental Drivers of Yellowfin Tuna (*Thunnus albacares*) Distribution Using SHapley Additive exPlanations (SHAP) Analysis. *Biology*, P.P 14(11), 1567. <https://doi.org/10.3390/biology14111567>
- Druon, J. N., Fromentin, J. M., et al. (2016).** Habitat suitability of the Atlantic bluefin tuna by size class: An ecological niche approach. *Progress in Oceanography*. P.P 142, 30–46. <https://doi.org/10.1016/j.pocean.2016.01.002>
- Effrosynidis, D., Tsikliras, A., et al. (2020).** Species distribution modelling via feature engineering and machine learning for pelagic fishes in the mediterranean sea. *Applied Sciences (Switzerland)*, P.P 10(24), 1–23. <https://doi.org/10.3390/app10248900>

Bibliographie

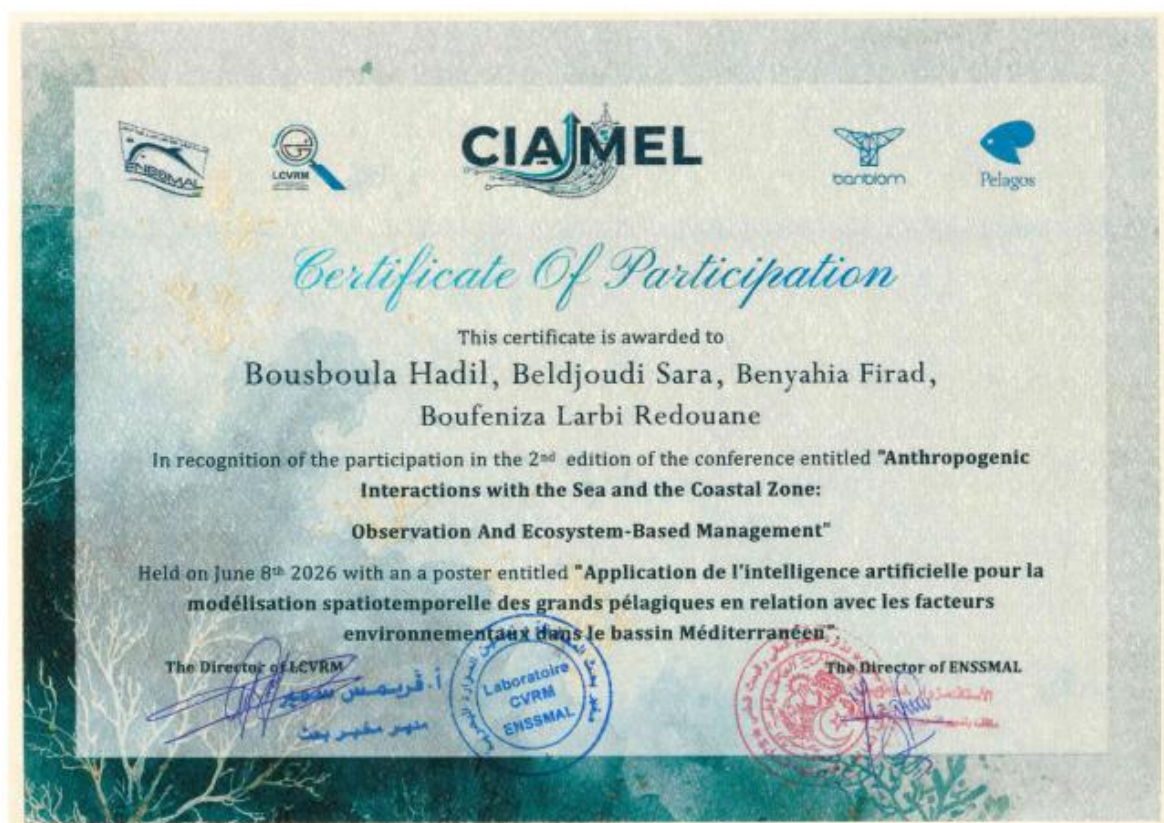
- Elith, J., & Leathwick, J. R. (2009).** Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, P.P 40, 677–697. <https://doi.org/10.1146/annurev.ecolsys.110308.120159>
- Fromentin, J. M., & Powers, J. E. (2005).** Atlantic bluefin tuna: Population dynamics, ecology, fisheries and management. In *Fish and Fisheries* (Vol. 6, Number 4, pp. 281–306). <https://doi.org/10.1111/j.1467-2979.2005.00197.x>
- Karp, M. A., Cimino, M., et al. (2025).** Applications of species distribution modeling and future needs to support marine resource management. *ICES Journal of Marine Science*, 82(3). <https://doi.org/10.1093/icesjms/fsaf024>
- Kühn, B., Cayetano, A., et al. (2025).** Machine Learning Applications for Fisheries—At Scales from Genomics to Ecosystems. In *Reviews in Fisheries Science and Aquaculture* (Vol. 33, Number 2, pp. 334–357). Taylor and Francis Ltd. <https://doi.org/10.1080/23308249.2024.2423189>
- Millot, C., & Taupier-Letage, I. (2005).** *Circulation in the Mediterranean Sea* (pp. 29–66). <https://doi.org/10.1007/b107143>
- Muhling, B. A., Brill, R., et al. (2017).** Projections of future habitat use by Atlantic bluefin tuna: Mechanistic vs. correlative distribution models. *ICES Journal of Marine Science*, 74(3), P.P 698–716. <https://doi.org/10.1093/icesjms/fsw215>
- Pan, S. J., & Yang, Q. (2010).** A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), P.P 1345–1359. <https://doi.org/10.1109/TKDE.2009.191>
- Precioso, D., Navarro-García, M., et al. (2021).** Tuna-AI: tuna biomass estimation with Machine Learning models trained on oceanography and echosounder FAD data. <https://doi.org/10.1016/j.fishres.2022.106263>
- Russo, S., Torri, M., et al. (2021).** Unveiling the Relationship Between Sea Surface Hydrographic Patterns and Tuna Larval Distribution in the Central Mediterranean Sea. *Frontiers in Marine Science*, 8. <https://doi.org/10.3389/fmars.2021.708775>
- Teo, S. L. H., Boustany, A. M., & Block, B. A. (2007).** Oceanographic preferences of Atlantic bluefin tuna, *Thunnus thynnus*, on their Gulf of Mexico breeding grounds. *Marine Biology*, 152(5), P.P 1105–1119. <https://doi.org/10.1007/s00227-007-0758-1>

Annexes

Annexes

Institution	Pays	Preuve (publication récente)	Année
IFREMER	France	"Machine learning prediction of BFT catches in Mediterranean" - Fisheries Research	2023
IEO-CSIC	Espagne	"Deep learning for tuna detection from remote sensing" - ICES J. Marine Science	2022
HCMR	Grèce	"AI models for pelagic fish distribution in Eastern Med" - Marine Ecology Progress Series	2023
University of Genoa	Italie	"Spatio-temporal modeling of tuna migration using ANN" - Ecological Modelling	2022
INSTM	Tunisie	"Data mining approaches for fisheries data in Central Med" - Mediterranean Marine Science	2021
INRH	Maroc	"Predictive modeling of pelagic resources" - African J. Marine Science	2022
CNR-ISMAR	Italie	"Satellite data and ML for fishery forecasting" - Remote Sensing	2023
METU	Turquie	"AI applications in Mediterranean fisheries management" - Turkish J. Fisheries	2022

Analyse des capacités de recherche en intelligence artificielle pour la gestion des pêches en Méditerranée



Annexes

Titre de l'article	Auteurs	Période des données (années)	Revue/SOURCE	Mots-clés	Sources de données	Zone d'étude	Modèles utilisés	Modèles comparés	Métriques	Thème / Application principale	Principaux résultats / Lacunes identifiées	Référence / Citation	DOI / Lien
Habitat suitability of the Atlantic bluefin tuna by size class: An ecological niche approach	Draon, J.-N.; Fromentin, J.-M.; Hanks, A.R.; Arribasbalga, H.; Damalas, D.; T'ic'ou, V.; Quikzo-Badri, G.; Ramirez, K.; Arregui, I.; Tserpes, G.; Reglero, P.; Deflorio, M.; Oray, I.; Karakulak, F.S.; Megalofonou, P.; Ceylan, T.; Grubišić, L.; MacKenzie, B.R.; Larkin, J.; Alfonso, P.; Addis, P.	1998-2009 (Gulf of Mexico); 2001-2005 (Baltic Islands, Mediterranean)	Progress in Oceanography	Athletic bluefin tuna.Larval habitat.Spawning habitat.Environmental drivers Gulf of Mexico,Mediterranean Sea,Temperature,Chlorophyll,a neural networks,PCA	Présences de thon rouge (pêcheries, marquage électronique), MODIS-Aqua (chlorophylle-a), Modèles océaniques MyOcean (SST, SSHa, SSCV)	Océan Atlantique Nord, Mer Méditerranée, Golfe du Mexique	Modélisation de niche écologique (ENM) - Analyse par clusters et enveloppes environnementales	ANN vs GLM	Distance des données de présence à l'habitat favorable	Prédiction de l'habitat potentiel d'alimentation et de ponte du thon rouge de l'Atlantique par classe de taille	Résultats : relier distribution/abondance et variables environnementales sur Océan Atlantique Nord, Mer Méditerranée, Golfe du Mexique via Modélisation de niche écologique (ENM) - Analyse par clusters et enveloppes environnementales (évaluation : Distance des données de présence à l'habitat favorable). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Draon et al. (2016)	https://doi.org/10.1016/j.pocean.2016.01.002
Differential larval growth strategies and trophodynamics of larval Atlantic bluefin tuna (Thunnus thynnus) from two discrete spawning grounds	Macia, E.; Quinamilla, J.M.; Gerard, T.; Alemany, F.; Sutton, T.; Garcia, A.; Larkin, J.T.; Laiz-Carrión, R.	2002-2012	Frontiers in Marine Science	otolith microstructure, general additive models, early life history, stable isotope analysis, postflexion stage, bluefin tuna, highly migratory species	Échantillonnage larvaire in situ (filets S-10 et Bongo), CTD, spectrométrie de masse des isotopes stables ($\delta^{15}N$, $\delta^{13}C$)	Golfe du Mexique, mer Méditerranée (mer des Baléares)	Modèles additifs généralisés (GAM), analyses de variance (ANCOVA, PERMANOVA), ellipses bayésiennes d'isotopes stables (SIBER)	Ecological Niche Model (ENM) / Habitat suitability model basé sur variables environnementales	Coefficient de variation (CV) pour l'âge, test t, test de Wilcoxon, AIC, variance expliquée, surface de l'ellipse standard (SEA)	Comparaison de la croissance larvaire de la trophodynamique du thon rouge de l'Atlantique entre deux zones de ponte	Résultats : relier distribution/abondance et variables environnementales sur Golfe du Mexique, mer Méditerranée (mer des Baléares) via Modèles additifs généralisés (GAM), analyses de variance (ANCOVA, PERMANOVA), ellipses bayésiennes d'isotopes stables (SIBER) (évaluation : Coefficient de variation (CV) pour l'âge, test t, test de Wilcoxon, AIC, variance expliquée, surface de l'ellipse standard (SEA)). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Généralisation à tout le bassin nécessite données harmonisées.	Macia et al. (2023)	https://doi.org/10.3389/fmars.2023.1233249
Surfacing and vertical behaviour of Atlantic bluefin tuna (Thunnus thynnus) in the Mediterranean Sea: implications for aerial surveys	Bauer, R. K.; Forget, F.; Fromentin, J.-M.; Capello, M.	2015-2016	ICES Journal of Marine Science	abundance index, archival tags, surface availability, thermal stratification, vertical behaviour	Pop-up satellite archival tags (miniPATs, Wildlife Computers); Sea surface temperature (SST) and bathymetry reference data.	Gulf of Lion, north-western Mediterranean Sea	GPE3 geolocation algorithm (gridded hidden Markov model); Kernel density estimation (kde2d); Statistical tests (Kruskal-Wallis, Pearson's correlation).	Modèles de géolocalisation GPE testés avec différentes vitesses (2.1 km/h, 3.1 km/h, 4.1 km/h)	Presence rates (%) in depth layers; Continuous residence time (CRT); Correlation coefficients.	Étude du comportement de surface et de la fidélité au site du thon rouge de l'Atlantique pour améliorer les indices d'abondance issus de relevés aériens.	Résultats : relier distribution/abondance et variables environnementales sur Gulf of Lion, north-western Mediterranean Sea via GPE3 geolocation algorithm (gridded hidden Markov model); Kernel density estimation (kde2d); Statistical tests (Kruskal-Wallis, Pearson's correlation). (évaluation : Presence rates (%) in depth layers; Continuous residence time (CRT); Correlation coefficients.). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Bauer et al. (2020)	https://doi.org/10.1093/icesjms/ksaa083
Pelagic seascape ecology for operational fisheries oceanography: modelling and predicting spawning distribution of Atlantic bluefin tuna in Western Mediterranean	Aharez-Berastegui, D.; Hidalgo, M.; Tugores, M. P.; Reglero, P.; Aparicio-Gonzalez, A.; Cinnelli, L.; Irujo, M.; Moure, B.; Pascual, A.; López-Jurado, J. L.; Garcia, A.; Rodriguez, J. M.; Tintoré, J.; Alemany, F.	2001-2005	ICES Journal of Marine Science	Balearic sea, ecosystem, ecosystem-based fisheries management, essential fish habitat, habitat forecast, Mediterranean, seascape metrics	Larval abundance from bongo net surveys (2001-2005); SST from Mediterranean Forecasting System (MFS); Kinetic energy from AVISO altimetry; Chlorophyll-a from multisatellite images (RS-CHL).	Balearic Sea, Western Mediterranean	Generalized Additive Models (GAMs) with negative binomial distribution; Spatial smoothing (moving window average).	Modèle de référence sans variables environnementales (M1).	Déviante expliquée (%); Akaike Information Criterion (AIC); Delta AIC (AAIC); Corrélation de Spearman.	Modélisation et prédiction de l'habitat de ponte du thon rouge de l'Atlantique à l'aide de métriques de paysage pélagique	Résultats : relier distribution/abondance et variables environnementales sur Balearic Sea, Western Mediterranean via Generalized Additive Models (GAMs) with negative binomial distribution; Spatial smoothing (moving window average). (évaluation : Déviante expliquée (%); Akaike Information Criterion (AIC); Delta AIC (AAIC); Corrélation de Spearman). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Généralisation à tout le bassin nécessite données harmonisées.	Aharez-Berastegui et al. (2016)	https://doi.org/10.1093/icesjms/ksw041
Projections of future habitat use by Atlantic bluefin tuna: mechanistic vs. correlative distribution models	Muhling, B. A.; Brill, R.; Larkin, J. T.; Roffler, M. A.; Lee, S.-K.; Lü, Y.; Muller-Karger, F.	1997-2014	ICES Journal of Marine Science	Athletic bluefin tuna, climate change, North Atlantic, species distribution models.	NOAA pelagic longline fishery logbook data; ICAT Task II database (catch rates); World Ocean Atlas (temperature, oxygen); Satellite data (Pathfinder, MODIS, SeaWiFS); HYCOM-NCODA GLBu0.08 reanalysis; GFDL-ESM2M Earth System Model output.	North Atlantic Ocean	Mechanistic Oxygen Balance Model (Metabolic/GAM, Gill Ventilation, Cardiac Transport); Correlative Species Distribution Model (Boosted Regression Tree - BRT).	Mechanistic model vs. Correlative SDM; Sensitivity analysis of SDM with constant variables.	t-test (significance of projected changes); Visual/model comparison with catch rate data; Model fit (GAM R ² , BRT variable importance).	Projection de l'habitat futur du thon rouge de l'Atlantique face au changement climatique.	Résultats : relier distribution/abondance et variables environnementales sur North Atlantic Ocean via Mechanistic Oxygen Balance Model (Metabolic/GAM, Gill Ventilation, Cardiac Transport); Correlative Species Distribution Model (Boosted Regression Tree - BRT). (évaluation : t-test (significance of projected changes); Visual/model comparison with catch rate data; Model fit (GAM R ² , BRT variable importance)). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Muhling et al. (2017)	https://doi.org/10.1093/icesjms/ksw215
Environmental characteristics of Atlantic bluefin tuna (Thunnus thynnus) larval habitat in the Gulf of Mexico and western Mediterranean Sea	Muhling, B.A.; Reglero, P.; Alemany, F.; Aharez-Berastegui, D.; Ingram, G.W.; Roffler, M.A.; Lalangui, G.; Cinnelli, L.	1998-2009 (Gulf of Mexico); 2001-2005 (Balearic Islands)	Marine Ecology Progress Series	Bluefin tuna - Spawning - Mediterranean Sea - Gulf of Mexico	Historical larval collections (SEAMAP, TUNIBAL projects), in situ CTD casts, remotely sensed satellite data (SeaWiFS, MODIS-Aqua), Aviso altimetry data, HYCOM-NCODA model outputs	Gulf of Mexico (northern), Western Mediterranean Sea (Balearic Islands)	Multi-layer perceptron artificial neural network (ANN), Principal Component Analysis (PCA)	PCA (Principal Component Analysis) + Neural Network models	Model sensitivity (~80%)	Modélisation de l'habitat larvaire du thon rouge de l'Atlantique	Résultats : relier distribution/abondance et variables environnementales sur Gulf of Mexico (northern), Western Mediterranean Sea (Balearic Islands) via Multi-layer perceptron artificial neural network (ANN), Principal Component Analysis (PCA) (évaluation : Model sensitivity (~80%)). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Muhling et al. (2013)	https://doi.org/10.3354/meps10397
Vertical distribution of Atlantic bluefin tuna Thunnus thynnus and bonito Sarda sarda larvae is related to temperature preference	Reglero, P.; Blanco, E.; Alemany, F.; Ferrá, C.; Aharez-Berastegui, D.; Ortega, A.; de la Gándara, F.; Aparicio-González, A.; Folkword, A.	2010-2016	Marine Ecology Progress Series	Thermocline - Vertical distribution - Temperature-dependent behaviour - Larval fish - Tuna - Bonito	Données d'échantillonnage vertical sur le terrain (filets multi-réseaux, MOCNESS), mesures de CTD in situ, expériences en colonnes d'eau contrôlées en laboratoire	Mer des Baléares, Méditerranée nord-occidentale	Modèles additifs généralisés (GAM), Modèles linéaires généralisés (GLM)	Generalized Additive Models (GAM)	Pourcentage de déviante expliquée, tests de signification statistique (p-adj)	Influence de la température sur la distribution verticale des larves de thon et de bonito	Résultats : relier distribution/abondance et variables environnementales sur Mer des Baléares, Méditerranée nord-occidentale via Modèles additifs généralisés (GAM), Modèles linéaires généralisés (GLM) (évaluation : Pourcentage de déviante expliquée, tests de signification statistique (p-adj)). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Généralisation à tout le bassin nécessite données harmonisées.	Reglero et al. (2018)	https://doi.org/10.3389/fmars.2021.708775
Modelling retention and dispersion mechanisms of bluefin tuna eggs and larvae in the northwest Mediterranean Sea	Mariani, P.; MacKenzie, B.R.; Iudicone, D.; Bozoz, A.	1999-2003	Progress in Oceanography	Bluefin tuna larvae; larval dispersion; hydrodynamic modelling; particle tracking; Mediterranean Sea; spawning habitat	Données de circulation océanique (modèle OPA), forçages atmosphériques (ECMWF), observations in situ d'œufs et de larves (campagnes TUNIBAL), données de dérivées	Mer des Baléares, Méditerranée nord-occidentale	Modèle individu-centré couplé biophysique Lagrangien, algorithme de suivi de particules avec marche aléatoire corrigée	Hydrodynamic circulation model + Lagrangian particle-tracking model	Comparaison qualitative des distributions modélisées/observées, indices d'aggrégation et de patchiness	Modélisation de la rétention et de la dispersion des premiers stades de vie du thon rouge	Résultats : relier distribution/abondance et variables environnementales sur Mer des Baléares, Méditerranée nord-occidentale via Modèle individu-centré couplé biophysique Lagrangien, algorithme de suivi de particules avec marche aléatoire corrigée (évaluation : Comparaison qualitative des distributions modélisées/observées, indices d'aggrégation et de patchiness). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Mariani et al. (2010)	https://doi.org/10.1016/j.pocean.2010.04.027

Annexes

Titre de l'article	Auteurs	Période des données (années)	Revue/SOURCE	Mots-clés	Sources de données	Zone d'étude	Modèles utilisés	Modèles comparés	Métriques	Thème / Application principale	Principaux résultats / Lacunes identifiées	Référence / Citation	DOI / Lien
Oceanographic preferences of Atlantic bluefin tuna, <i>Thunnus thynnus</i> , on their Gulf of Mexico breeding grounds	Teo, S.L.H.; Boustany, A.M.; Block, B.A.	1996–2005	Marine Biology	N/A (Inferred: Atlantic bluefin tuna, habitat preference, spawning, Gulf of Mexico, electronic tagging)	Implantable archival tags, pop-up satellite archival tags (PAT), AVHRR Pathfinder SST, SeaWiFSMODIS chlorophyll, AVISO SSHA & geostrophic velocity, ERS-QuikSCAT wind, CLS surface currents	Golfe du Mexique	Indice de préférence de Chersson, modèle de choix discret (régression logit multinomial), fonction de sélection des ressources (RSF)	Modèles nuls générés par la méthode de Monte Carlo	Valeur P (tests de Monte Carlo), sensibilité de la RSF, coefficients du modèle de choix discret	Identification des préférences d'habitat océanographique pour la reproduction du thon rouge de l'Atlantique	Résultats : relie distribution/abondance et variables environnementales sur Golfe du Mexique via Indice de préférence de Chersson, modèle de choix discret (régression logit multinomial), fonction de sélection des ressources (RSF) (évaluation : Valeur P (tests de Monte Carlo), sensibilité de la RSF, coefficients du modèle de choix discret). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Teo et al. (2007)	https://doi.org/10.1007/s00227-007-0739-2
Projecting climate change impacts on the distribution of the most captured tuna species on Atlantic waters	Silva, P.M.; Borges, F.O.; Pita, C.; Costa, J.L.; Teixeira, C.M.	2010–2020 (baseline) ; projections 2030 et 2050	Biodiversity and Conservation	Species distribution models - Fisheries - Sustainability - Atlantic	GBIF, ICCAT, Bio-ORACLE v3.0 (CMIP6 couches environnementales - SST, salinité, oxygène, productivité primaire, pente)	Océan Atlantique et Mer Méditerranée	MaxEnt, modèle de distribution d'espèces (SDM), package R 'megaSDM'	Species Distribution Model (SDM) – MaxEnt algorithm	AUC (Aire sous la courbe)	Projection de l'impact du changement climatique sur l'habitat et la distribution de cinq espèces de thon de l'Atlantique (2030, 2050)	Résultats : relie distribution/abondance et variables environnementales sur Océan Atlantique et Mer Méditerranée via MaxEnt, modèle de distribution d'espèces (SDM), package R 'megaSDM' (évaluation : AUC (Aire sous la courbe)). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Silva et al. (2025)	https://doi.org/10.1007/s10531-025-03121-x
Tuna-At: Tuna biomass estimation with Machine Learning models trained on oceanography and echosounder FAD data	Precioso, D.; Navarro-García, M.; Gavira-O'Neill, S.; Torre-Barrán, A.; Gordo, D.; Gallego-Alcalá, V.; Gómez-Ullate, D.	2018–2020	Preprint	Tunas - Direct abundance indicator - Echo-sounder buoys Fish aggregating devices - Purse seiner	FAD logbooks (AGAC, 2018–2020), données d'écho-sondeur de bouées Satellit (ISL -, SLX -, ISD+) 1, données océanographiques Copernicus Marine Service (CMEMS)	Océans Atlantique, Indien, Pacifique (tropical et sub-tropical)	Gradient Boosting (GB), Random Forest (RF), XGBoost (XGB), Régression Logistique (LR), Elastic Net (ENet)	Modèle de référence (baseline) basé sur des règles d'agrégation de la biomasse acoustique	AUC ROC, F1-score (pondéré), MAE	Estimation de la biomasse de thon sous les DCPs à l'aide du ML, combinant données acoustiques, océanographiques et de captures	Résultats : relie distribution/abondance et variables environnementales sur Océans Atlantique, Indien, Pacifique (tropical et sub-tropical) via Gradient Boosting (GB), Random Forest (RF), XGBoost (XGB), Régression Logistique (LR), Elastic Net (ENet) (évaluation : AUC ROC, F1-score (pondéré), MAE). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Precioso et al. (N/A)	https://doi.org/10.48550/arXiv.2109.06732
Environmental Conditions along Tuna Larval Dispersion: Insights on the Spawning Habitat and Impact on Their Development Stages	Russo, S.; Torri, M.; Patti, B.; Musco, M.; Masullo, T.; Di Natale, M.V.; Sara, G.; Cuttitta, A.	2010–2016	Water	<i>Thunnus thynnus</i> ; Axiolus rochei; <i>Thunnus alalunga</i> ; Mediterranean Sea; backward trajectories; Lagrangian simulations; spawning habitat; larval habitat; tuna	In-situ échantillonnage ichtyoplanktonique (2010–2016), Données alimentaires CMEMS, Données vent NOAA/NCEP, SST et Chl-a Copernicus Marine Service (résolution 0.01° et 1 km), ETOPO1 bathymétrie	Détroit de Sicile, Méditerranée centrale	Simulations Lagrangiennes rétrogrades (GNOME), Tests statistiques non-paramétriques (Kruskal-Wallis, Wilcoxon)	Random Forest (RF) ; Gradient Boosting Machines (GBM) ; Neural Networks (NN)	RMSE (Root Mean Square Error) ; MAE (Mean Absolute Error) ; R² (coefficient de détermination)	Caractérisation des habitats de ponte et d'influence des conditions environnementales (SST, Chl-a) sur le développement larvaire de trois espèces de thon.	Résultats : relie distribution/abondance et variables environnementales sur Détroit de Sicile, Méditerranée centrale via Simulations Lagrangiennes rétrogrades (GNOME), Tests statistiques non-paramétriques (Kruskal-Wallis, Wilcoxon) (évaluation : N/A). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Russo et al. (2022)	https://doi.org/10.3390/w14101568
Geographically and environmentally driven spawning distributions of tuna species in the western Mediterranean Sea	Reglero, P.; Canneli, L.; Alvarez-Bernastegui, D.; Balbià, R.; López-Jurado, J.L.; Alemany, F.	2001–2005	Marine Ecology Progress Series	spawning habitat; tuna larvae; generalized additive models; Mediterranean; environmental drivers	Campagnes larvaires (Bongo nets) + CTD; salinité; vitesse géostrophique; résidus de température (couche mélange)	Méditerranée occidentale (Iles Baléares)	GAM (modèle géographique + modèle hydrographique)	Comparaison modèle géographique vs hydrographique	R², GCV, effets partiels des covariables (splines)	Distribution spatio-temporelle des zones de ponte (larves) et rôle des facteurs environnementaux vs géographiques	Résultats : relie distribution/abondance et variables environnementales sur Méditerranée occidentale (Iles Baléares) via GAM (modèle géographique + modèle hydrographique) (évaluation : R², GCV, effets partiels des covariables (splines)). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Généralisation à tout le bassin nécessite données harmonisées.	Reglero et al. (2012)	https://doi.org/10.3354/meps9800
A novel framework for spatio-temporal prediction of environmental data using deep learning	Federico Amato; Fabian Guignard; Sylvain Robert; Mikhail Kanevski	Données simulées + données environnementales réelles (période non spécifiée)	Scientific Reports (2020)	spatio-temporal; deep learning; interpolation; environmental data	Données environnementales (stations/points irréguliers) + 2 études de cas (simulé + réel)	Études de cas (local/régional)	Framework DL spatio-temporel (base fonctions + coefficients spatiaux)	Deep Feedforward Neural Network ; PCA / EOF decomposition ; méthodes géostatistiques (kriging)	RMSE ; analyse de variogramme spatio-temporel	Prédiction spatio-temporelle de données environnementales	Résultats : relie distribution/abondance et variables environnementales sur Études de cas (local/régional) via Framework DL spatio-temporel (base fonctions + coefficients spatiaux). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Amato et al., 2020	10.1038/s41598-020-79148-7
Machine Learning Applications for Fisheries—At Scales from Genomics to Ecosystems	Bernhard Kühn; Arjay Cayetano; Jennifer I. Flucham; Hassan Moustahfid; Maria Sokolova; Neda Trifunova; Jordan T. Watson; Jose A. Fernandes-Salvador; Laura Usatillo	Revue scientifique – plusieurs études avec périodes de données différentes	Reviews in Fisheries Science & Aquaculture (2024)	machine learning; fisheries; monitoring; management; review	Revue de littérature	Globale	Panorama d'approches ML (supervisé/ non supervisé etc.)	Random Forest (RF) ; Convolutional Neural Networks (CNN) ; Artificial Neural Networks (ANN) ; Support Vector Machines (SVM) ; Naïve Bayes ; Decision Trees ; Renforcement Learning	Accuracy ; Precision ; Recall ; F1-score ; RMSE	Revue des applications ML en sciences halieutiques	Résultats : synthèse des approches et usages liés à « Revue des applications ML en sciences halieutiques ». Lacunes : pas de cas d'application unique avec données originales ; met surtout en avant des besoins (données, validation externe, transférabilité, interprétabilité).	Kühn et al., 2024	10.1080/23308249.2024.2423189
Joint dynamic species distribution models: a tool for community ordination and spatiotemporal monitoring	James T. Thorson; James N. Jewell; Elise A. Larsen; Leslie Rice; Mark D. Schweerdt; Cody Sawwaki; Elise F. Zipkin	1982–2014 (33 ans, Bering Sea survey) + 2010 (Ohio butterfly dataset)	Global Ecology and Biogeography (2016)	joint SDM; dynamic SDM; geostatistics; spatiotemporal; community	2 études de cas (Bering Sea survey annuelle; USA survey saisonnier papillons)	Bering Sea + USA continental	JDSMD (dynamic factor analysis + géostatistique)	Joint Dynamic Species Distribution Model (JDSMD) ; Spatial Dynamic Factor Analysis (SDFA) ; Species Distribution Models (SDM)	Log-likelihood ; model fit comparison ; spatial autocorrelation diagnostics	Modèles SDM joints et dynamiques (monitoring spatio-temporel)	Résultats : relie distribution/abondance et variables environnementales sur Bering Sea + USA continental via JDSMD (dynamic factor analysis + géostatistique). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Thorson et al., 2016	10.1111/geb.12464

Annexes

Titre de l'article	Auteurs	Période des données (années)	Revue/SOURCE	Mots-clés	Sources de données	Zone d'étude	Modèles utilisés	Modèles comparés	Métriques	Thème / Application principale	Principaux résultats / Lacunes identifiées	Référence / Citation	DOI / Lien
Machine Learning Applications for Fisheries—At Scales from Genomics to Ecosystems	Bernhard Kuhn; Arjy Gayetano; Jennifer I. Fincham; Hassan Moustahfid; Maria Sokolova; Neda Trifunova; Jordan T. Watson; Jose A. Fernandes-Sabador; Laura Usakalo	Revue scientifique – plusieurs études avec périodes de données différentes	Reviews in Fisheries Science & Aquaculture (2024)	machine learning; fisheries; monitoring; management; review	Revue de littérature	Global	Panorama d'approches ML (supervisé/ non supervisé etc.)	Random Forest (RF) ; Convolutional Neural Networks (CNN) ; Artificial Neural Networks (ANN) ; Support Vector Machines (SVM) ; Naïve Bayes ; Decision Trees ; Reinforcement Learning	Accuracy ; Precision ; Recall ; F1-score ; RMSE	Revue des applications ML en sciences halieutiques	Résultats : synthèse des approches et usages liés à « Revue des applications ML en sciences halieutiques ». Lacunes : pas de cas d'application unique avec données originales ; met surtout en avant des besoins (données, validation externe, transférabilité, interprétabilité).	Kuhn et al., 2024	10.1080/23308240.2024.2423189
Joint dynamic species distribution models: a tool for community ordination and spatiotemporal monitoring	James T. Thorson; James N. Ianelli; Elise A. Larsen; Leslie Rice; Mark D. Scheuerell; Cody Szawalski; Elke F. Ziplan	1982–2014 (33 ans, Bering Sea survey) + 2010 (Ohio butterfly dataset)	Global Ecology and Biogeography (2016)	joint SDM; dynamic SDM; geostatistics; spatiotemporal; community	2 études de cas (Bering Sea survey annuel; USA survey saumonier papillons)	Bering Sea + USA continental	JDSDM (dynamic factor analysis + géostatistique)	Joint Dynamic Species Distribution Model (JDSDM) ; Spatial Dynamic Factor Analysis (SDFA) ; Species Distribution Models (SDM)	Log-likelihood ; model fit comparison ; spatial autocorrelation diagnostics	Modèles SDM joints et dynamiques (monitoring spatio-temporel)	Résultats : relie distribution/abondance et variables environnementales sur Bering Sea + USA continental via JDSDM (dynamic factor analysis + géostatistique). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Thorson et al., 2016	10.1111/geb.12464
A systematic review of spatial habitat associations and modeling of marine fish distribution: A guide to predictors, methods, and knowledge gaps	Bradley A. Pickens; Rachel Carroll; Michael J. Schirripa; Francesca Forrestal; Kevin D. Friedland; J. Christopher Taylor	2007–2018	PLOS ONE (2021)	SDM; marine fish; predictors; methods; knowledge gaps; review	Revue de 225 articles (2007–2018)	Global	Analyse de littérature + comparaisons méthodes/prédicteurs	Generalized Additive Model (GAM) ; Generalized Linear Model (GLM) ; Maxent ; Habitat Suitability Index ; Machine Learning models (Random Forest, Boosted Regression Trees, Artificial Neural Networks, Support Vector Machines, MARS) ; Geostatistical models (Kriging) ; Bayesian models ; Quantile regression ; Individual-based models	Multinomial logistic regression ; Linear Discriminant Analysis (LDA) ; Frequency analysis ; model comparison statistics	Revue SDM poissons marins (prédicteurs, méthodes, lacunes)	Résultats : synthèse des approches et usages liés à « Revue SDM poissons marins (prédicteurs, méthodes, lacunes) ». Lacunes : pas de cas d'application unique avec données originales ; met surtout en avant des besoins (données, validation externe, transférabilité, interprétabilité).	Pickens et al., 2021	10.1371/journal.pone.0251818
Spatio-temporal model and machine learning method reveal patterns and processes of migration under climate change	Yuki Kanamori; Toshikazu Yano; Hiroshi Okamura; Yuta Yagi	1972–2019	Journal of Biogeography (2024)	spatio-temporal model; machine learning; migration; xAI; SST	Séries présence/absence ~5 décennies + covariables (SST, profondeur, champ magnétique, productivité)	Pacifique Nord-Ouest (WNP)	Barrier model (spatio-temporel) + gradient boosting + xAI	Barrier spatio-temporal model ; Gradient Boosting Machine (machine learning) ; Linear regression model	RMSE (Root Mean Squared Error) ; Permutation Feature Importance (PFI) ; Grouped Permutation Feature Importance (GPFI) ; Cross-validation	Migration sous changement climatique: séparation effets spatiaux/ temporels/ spatio-temporels	Résultats : relie distribution/abondance et variables environnementales sur Pacifique Nord-Ouest (WNP) via Barrier model (spatio-temporel) + gradient boosting + xAI. Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Kanamori et al., 2024	10.1111/jbi.14595
Applications of species distribution modeling and future needs to support marine resource management	Melissa A. Karp et al.	Variable selon les études et les exemples (plusieurs séries temporelles, ex. 1982–2019 et 2021–2023 pour certaines analyses de poisson)	ICES Journal of Marine Science (2025)	SDM; fisheries management; marine spatial planning; climate change	Revue / synthèse d'applications SDM	Global	Panorama SDM (GLM/GAM, ML, modèles spatio-temporels)	Species Distribution Models (SDMs) ; Generalized Linear Models (GLM) ; Generalized Additive Models (GAM) ; Maximum Entropy (MaxEnt) ; Boosted Regression Trees ; Spatio-temporal models (VAST) ; Machine Learning algorithms	Coefficient of variation (CV) ; Indices of abundance standardization ; CPUE standardization ; Model comparison statistics	Besoins futurs et usages SDM en gestion des ressources marines	Résultats : relie distribution/abondance et variables environnementales sur Global via Panorama SDM (GLM/GAM, ML, modèles spatio-temporels). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Karp et al., 2025	10.1093/icesjms/fsaf024
Spatio-temporal species distribution models reveal dynamic indicators for ecosystem-based fisheries management	J. I. Badger; S. I. Large; J. T. Thorson	1963–2019	ICES Journal of Marine Science (2023)	VAST; ecosystem indicators; spatio-temporal; IEA	Données de survey (NE US Shelf Ecosystem)	Northeast US Shelf (Georges Bank, Gulf of Maine, Mid-Atlantic Bight)	Modèles spatio-temporels multivariés (VAST / VAR spatio-temporel)	Indicateurs agrégés vs spatio-temporels	Indicateurs/variabilité (selon cas d'étude)	Indicateurs dynamiques pour gestion écosystémique	Résultats : relie distribution/abondance et variables environnementales sur Northeast US Shelf (Georges Bank, Gulf of Maine, Mid-Atlantic Bight) via Modèles spatio-temporels multivariés (VAST / VAR spatio-temporel) (évaluation : Indicateurs/variabilité (selon cas d'étude)). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Badger et al., 2023	10.1093/icesjms/fsad123
Species Distribution Models: Ecological Explanation and Prediction Across Space and Time	Jane Elith; John R. Leathwick	Variable selon les études utilisées dans la revue (pas de période unique – article de synthèse)	Annual Review of Ecology, Evolution, and Systematics (2009)	SDM; prediction; presence-only; scale; uncertainty	Revue	Global	Synthèse méthodes SDM	Generalized Linear Models (GLM) ; Generalized Additive Models (GAM) ; Maximum Entropy (MaxEnt) ; Resource Selection Functions (RSF) ; Regression models ; Spatial statistical models ; Ecological niche models	Model fit statistics ; Residual analysis ; Predictive performance ; Validation with test data	Revue des SDM (explication & prédiction espace/temps)	Résultats : synthèse des approches et usages liés à « Revue des SDM (explication & prédiction espace/temps) ». Lacunes : pas de cas d'application unique avec données originales ; met surtout en avant des besoins (données, validation externe, transférabilité, interprétabilité).	Elith & Leathwick, 2009	10.1146/annurev.ecolsys.110308.120159
Temporal analysis of the distribution of thynnus thynnus in relation to environmental factors along the Algerian coast	Find Benyahia; Bachari Nour El Islam; Khalid Amarouche; Ghoulam Allah Kames; Lamine Salim	2010–2019	Egyptian Journal of Aquatic Biology & Fisheries (2021)	Thunnus thynnus; chlorophyll-a; oxygen; salinity; temperature; Algerian coast	Données halieutiques géo/étiérées + données océanographiques téké/détection	Côte algérienne	Régression linéaire multiple	Régression linéaire multiple ; ANOVA ; PCA	R ² / variance expliquée	Déterminants environnementaux de la distribution temporelle	Résultats : relie distributions/abondance et variables environnementales sur Côte algérienne via Régression linéaire multiple (évaluation : R ² / variance expliquée). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Benyahia et al., 2021	https://doi.org/10.1080/15641260701484572

Annexes

Titre de l'article	Auteurs	Période des données (années)	Revue/Source	Mots-clés	Sources de données	Zone d'étude	Modèles utilisés	Modèles comparés	Métriques	Thème / Application principale	Principaux résultats / Lacunes identifiées	Référence / Citation	DOI / Lien
Temporal analysis of the distribution of thynnus thynnus in relation to environmental factors along the Algerian coast	Frad Benyahia; Bachari Nour El Islam; Khalid Amarouche; Ghoulam Allah Kames; Lamine Salm	2010–2019	Egyptian Journal of Aquatic Biology & Fisheries (2021)	Thunnus thynnus; chlorophyll-a; oxygen; salinity; temperature; Algerian coast	Données halieutiques géoréférences + données océanographiques télédétection	Côte algérienne	Régression linéaire multiple	Régression linéaire multiple ; ANOVA ; PCA	R ² / variance expliquée	Déterminants environnementaux de la distribution temporelle	Résultats : relie distribution/abondance et variables environnementales sur Côte algérienne via Régression linéaire multiple (évaluation : R ² / variance expliquée). Lacunes : limites des hypothèses linéaires/paramétriques et faible capture des non-linéarités complexes. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Benyahia et al., 2021	https://doi.org/10.1080/15641260701484572
Deep Learning for the Deep Blue: A Systematic Review of AI in Marine and Freshwater Biodiversity Monitoring	Suresh Palarimath	2010–2025	FishTaxa (2025)	deep learning; biodiversity monitoring; PRISMA; XAI; review	Revue systématique (PRISMA 2020)	Global	Synthèse DL/ML (vision, PAM, cartographie prédictive)	Generalized Additive Models (GAM) – présence/absence et abondance	Densité des larves (larvae/m ³) ; CPUE (Capture Per Unit Area) ; Probabilité de présence ; Abondance des larves ; Écart-type (SD) ; Fréquence d'occurrence.	Revue AI pour monitoring biodiversité aquatique	Résultats : synthèse des approches et usages liés à « Revue AI pour monitoring biodiversité aquatique ». Lacunes : pas de cas d'application unique avec données originales ; met surtout en avant des besoins (données, validation externe, transférabilité, interprétabilité).	Palarimath, 2025	https://doi.org/10.3389/fmars.2021.708775
Species Distribution Modelling via Feature Engineering and Machine Learning for Pelagic Fishes in the Mediterranean Sea	Dimitrios Effrosynidis; Athanasios Tskiklis; Avi Arampatzis; Georgios Sylaios	10 ans (mean-monthly)	Applied Sciences (2020)	SDM; pelagic fish; feature engineering; XGBoost; Mediterranean	AquaMaps (occurrence) + CMEMS/EMODnet (environnement)	Mer Méditerranée (bassin complet)	XGBoost (final) + feature engineering/selection	10 algorithmes ML (testés)	Performance par espèce (amélioration via sélection)	Cartes de probabilité d'occurrence à haute résolution	Résultats : relie distribution/abondance et variables environnementales sur Mer Méditerranée (bassin complet) via XGBoost (final) + feature engineering/selection (évaluation : Performance par espèce (amélioration via sélection)). Lacunes : nécessité d'interprétabilité, gestion des biais (effort d'échantillonnage, présence-seulement) et validation externe. Généralisation à tout le bassin nécessite données harmonisées.	Effrosynidis et al., 2020	10.3390/app10248900
Unveiling the Relationship Between Sea Surface Hydrographic Patterns and Tuna Larval Distribution in the Central Mediterranean Sea	Stefania Russo; Marco Torri; Bernardo Patti; Patricia Reglero; Diego Alvarez-Berastegui; Angela Cuttitta; Gianluca Sara	2010–2016	Frontiers in Marine Science (2021)	tuna larvae; hydrography; SST; salinity; Mediterranean	7 surveys annuels ichthyoplankton + paramètres in situ	Méditerranée centrale (Déroit de Sicile)	Modèles présence/absence & abondance (statistiques)	VAST (Vector Autoregressive Spatio-Temporal Model) vs CPUE standardisation models	CPUE (catch per unit effort), AIC (Akaike Information Criterion), indices d'abondance relative, centre de gravité spatial, aire d'occupation	Relations environnementales–distribution larvaire (ABT, bullet tuna, albacore)	Résultats : relie distribution/abondance et variables environnementales sur Méditerranée centrale (Déroit de Sicile) via Modèles présence/absence & abondance (statistiques). Lacunes : manque d'intégration multi-sources et de dynamique spatio-temporelle plus complète. Généralisation à tout le bassin nécessite données harmonisées.	Russo et al., 2021	10.3389/fmars.2021.708775
Spatio-temporal dynamics of bluefin tuna (Thunnus thynnus) in US waters of the northwest Atlantic	Alexander Hansell; Sarah Becker; Steven X. Cadrin; Matt Lauretta; John Walter; Lisa Kerr	1993–2020	Fisheries Research (Elsevier, 2022)	bluefin tuna; VAST; spatio-temporal; SST; distribution shift	Données pêcheries/indices + covariables (SST, profondeur, AMO, biomasse proies)	Eaux US Atlantique Nord-Ouest	VAST (vector auto-regressive spatio-temporal)	Indices standardisés classiques vs indices spatio-temporels	Indices d'abondance; centre de gravité; aire occupée	Détecter décalages spatio-temporels et produire indices robustes	Résultats : relie distribution/abondance et variables environnementales sur Eaux US Atlantique Nord-Ouest via VAST (vector auto-regressive spatio-temporal) (évaluation : indices d'abondance; centre de gravité; aire occupée). Lacunes : dépend fortement de la couverture des données et des covariables ; besoin de tests de robustesse et validation externe. Applicabilité directe à la Méditerranée à discuter si la zone diffère.	Hansell et al., 2022	https://doi.org/10.1016/j.fishres.2022.106208
Comparison between environmental characteristics of larval bluefin tuna (Thunnus thynnus) habitat in the Gulf of Mexico and western Mediterranean Sea	Muhling, B. A.; Reglero, P.; Ciannelli, L.; Alvarez-Berastegui, D.; Alemany, F.; Lamkin, J. T.; Roffer, M. A.	1998–2009 (GOM); 2001–2005 (W. Mediterranean)	Marine Ecology Progress Series	Bluefin tuna; Larvae; Spawning habitat; Gulf of Mexico; Western Mediterranean; Environmental drivers; Neural networks; PCA	SEAMAP larval surveys; Mediterranean larval surveys; CTD data; MODIS & SeaWiFS chlorophyll; Geostrophic velocity	Northern Gulf of Mexico; Western Mediterranean (Balearic Islands)	Principal Component Analysis (PCA); Artificial Neural Networks (Multilayer Perceptron)	Regional habitat models comparison (GOM vs Mediterranean)	Probability of occurrence; Sensitivity >80%; R ² relationships	Comparison of spawning habitat characteristics between Atlantic bluefin tuna spawning grounds	Larvae occur in warm (24–28°C), low-chlorophyll offshore waters. In the Gulf of Mexico, larvae avoid Loop Current waters. In the Mediterranean, larvae are concentrated near water-mass boundaries. Temperature strongly controls spawning timing. Ecological survival mechanisms and recruitment processes remain unclear; model predicts presence not abundance.	Muhling et al., 2013	https://doi.org/10.3354/meps10397