

Work Package 4

Caractérisation de l'effort de pêche et standardisation de la CPUE

(Marta Quinzan, Florian Quemper, Paz Sampedro)

- ✓ Nettoyage et traitement des données VMS et calcul de l'effort de pêche (h) pour chaque stock cible
- ✓ Validation et correction à l'aide des données d'effort des observateurs embarqués (OAB)
- ✓ Standardisation de la CPUE pour chaque stock cible

[illegible]

Aperçu du processus de standardisation des indices d'abondance commercial (CPUEs)

Qu'est-ce que la CPUE et pourquoi c'est important ?



CPUE: La capture par unité d'effort (CPUE), mesure la quantité de poisson capturée par rapport à l'effort de pêche, servant d'indicateur de l'abondance relative des stocks dans le domaine des sciences halieutiques.

$$CPUE = \frac{C}{E} = q \cdot N$$

C = capture (kg ou n)	q = capturabilité
E = effort	N = abondance

Importance: Les indices d'abondance basés sur les prises par unité d'effort constituent des éléments importants de nombreuses évaluations des stocks, en particulier lorsque des études indépendantes de la pêche ne sont pas disponibles.

L'idée de base - et le problème

- Qu'est-ce que la **CPUE nominale** (capture brute/effort)

Assumptions:

1. La CPUE est proportionnelle à l'abondance
2. La capturabilité ne change pas systématiquement au fil du temps
3. La proportion de la population (taille, âge, sexe, etc.) est connue et ne change pas systématiquement au fil du temps

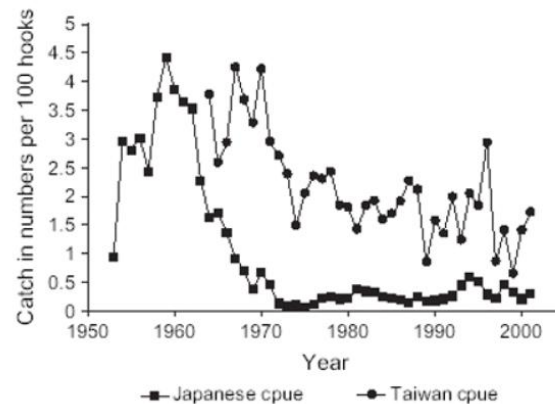
- Pourquoi CPUE nominale $\neq$ abondance

Le **hypothèse** of q constant est largement **violée** dans les pêcheries réelles, nécessitant une standardisation.

Facteurs Clés Affectant le CPUE



- l'engin et le type de bateau
- la zone et la saison
- les changements technologiques
- les changements économiques ou réglementaires (fermetures partielles, quotas de capture, limites de taille)
- l'espèce réellement ciblée dans les pêcheries multispécifiques

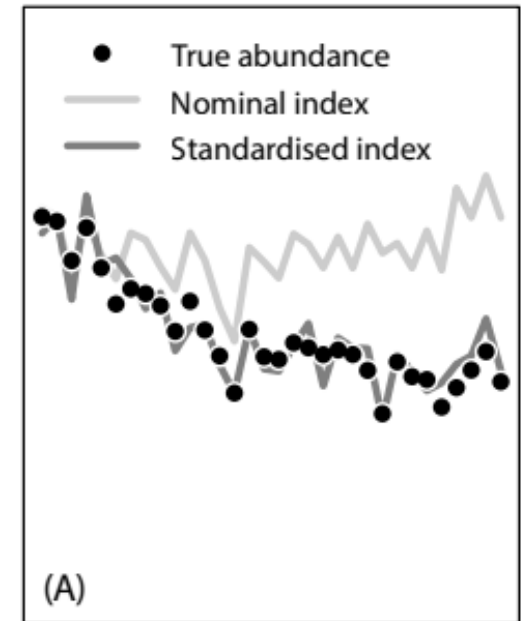


*Albacore in Pacific Ocean
(Myers and Worm, 2003)*

Ceci est essentiel car les modèles d'évaluation des stocks ont besoin d'un indice qui reflète les poissons et non les choix des pêcheurs.

Standardisation des CPUE

- Définition : est le processus statistique qui élimine les biais, les effets des facteurs non liés à l'abondance, du **CPUE nominale**
- Objectif : obtenir un indice proportionnel à l'abondance réelle
- De la CPUE nominale à la CPUE standardisée



Nettoyer les données de CPUE pour que la tendance au fil des années reflète au maximum l'abondance des poissons, et pas tout le reste.

Étapes de la standardisation

Partie 1 Collecte et évaluation de la qualité des données

Partie 2 Sélection et exploration des variables

Partie 3 Développement et sélection du modèle

Partie 4 Calcul de l'indice

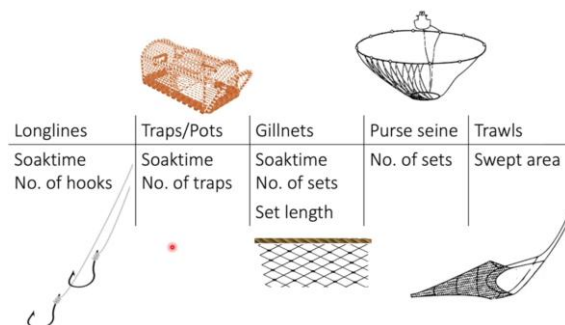
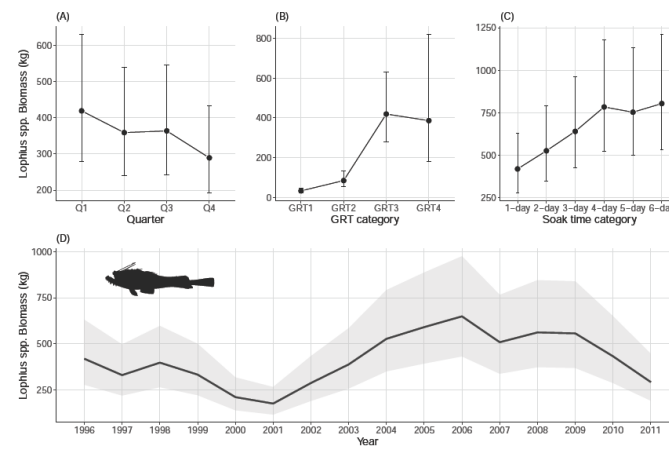
Partie 1 : Préparation des Données

1.Compilation et validation des données

2. Analyse exploratoire des données

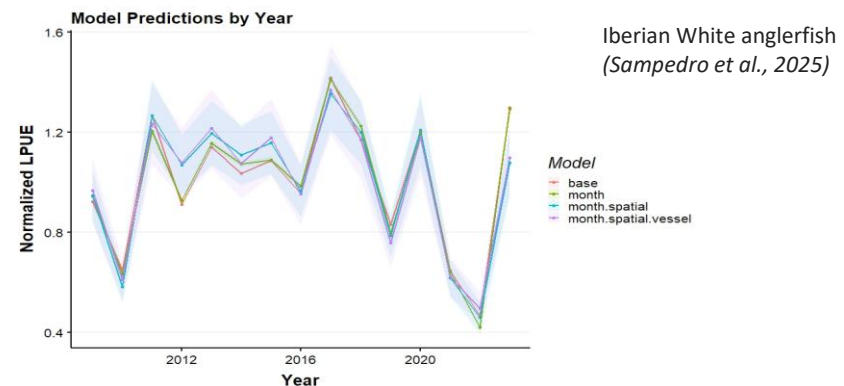
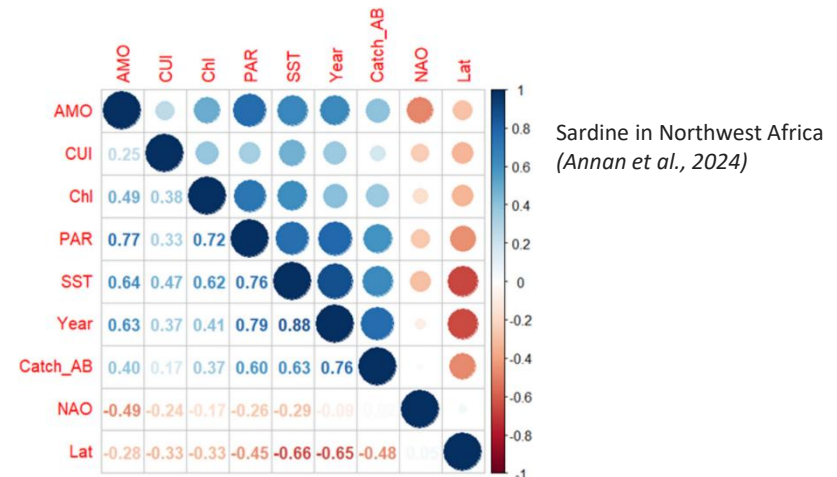
3.Identifier les valeurs aberrantes et les problèmes de qualité

4.Définir les métriques d'effort (kw*jours, heures de pêche, marées)



Partie 2 : Sélection des Variables

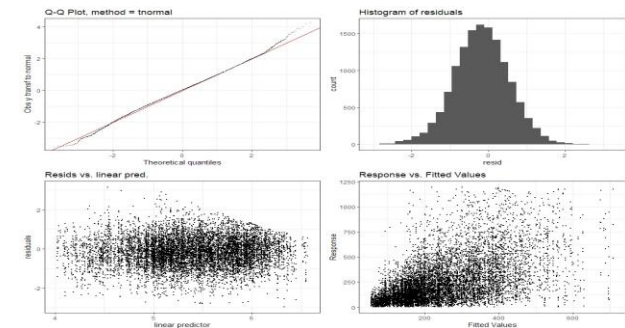
1. Identifier les variables explicatives candidates
2. Vérifier la colinéarité
3. Considérer la pertinence biologique
4. Équilibre entre complexité et parcimonie



Partie 3 : Développement du Modèle

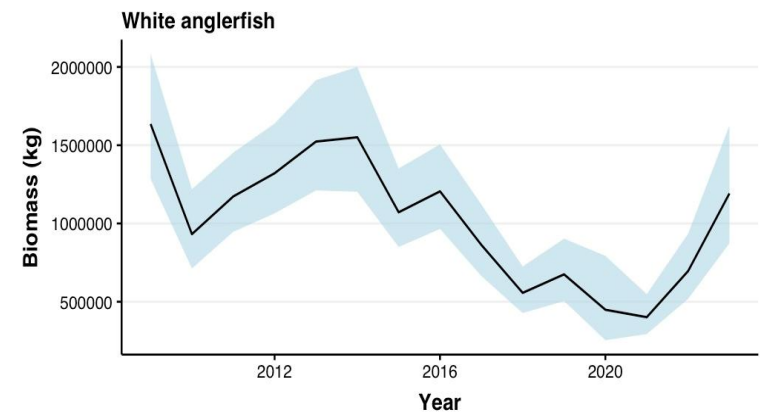
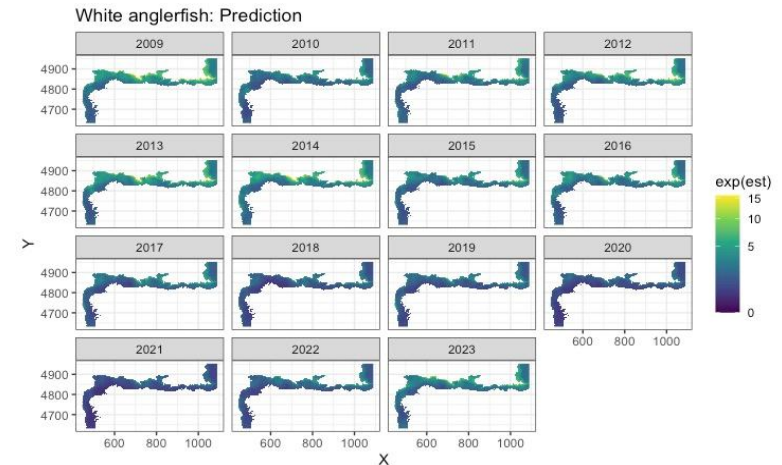
1. Choisir la distribution d'erreur appropriée (Gaussienne, Gamma, Binomiale Négative)
2. Sélectionner la fonction de lien (log, identité)
3. Ajuster les modèles candidats
4. Vérifier les résidus et les hypothèses du modèle

Model	AIC	Deviance	Scale (phi)
Tweedie	9278	5121	1.62
Gamma	5945	915	0.65



Partie 4 : Standardisation

1. Prédire les CPUE sur une grille spatio-temporelle standard
2. Fixer les variables d'efficacité de pêche à des valeurs standard
3. Calculer les indices annuels en moyennant les prédictions
4. Estimer l'incertitude (intervalles de confiance)



Aperçu des Méthodes Statistiques



Modèle	Hypothèses	Forces pour la standardisation CPUE	Faiblesses	Coût computationnel	R library
GLM	Prédicteurs linéaires, effets fixes uniquement (ex. Poisson, binomial négatif, lognormal)	Simple, rapide ; bon pour covariables basiques (année, engin, zone)	Linéarité paramétrique ; ignore effets aléatoires/spatiaux	Faible	glm()
GAM	Lissage non linéaires (splines/produits tensoriels) pour effets fixes	Gère effets env. non linéaires (SST, profondeur) ; meilleur ajustement que GLM	Effets fixes seulement ; pas de structure spatiale/aléatoire ; risque surajustement	Moyen	mgcv::gam(), bam()
GLMM	Effets fixes + aléatoires (ex. ID navire, année×zone comme intercepts/pentes)	Modélise hétérogénéité/clustering ; données hiérarchiques	Pas de champs spatio-temporels explicites ; indépendance conditionnelle	Moyen-Élevé	lme4::glmer(), glmmTMB
VAST	Delta-GLMM + analyse factorielle pour spatio-temporel multivarié ; TMB	Multivarié (multi-espèces/taille) ; prédiction surface densité ; usage écosystème mature	Courbe apprentissage raide (fichiers contrôle) ; problèmes reproductibilité ; lourd univarié	Élevé (évolutif TMB)	VAST
sdmTMB	GLMM + champs aléatoires spatio-temporels SPDE (approx. GMRF) ; backend TMB	Modélisation spatio-temporelle rapide ; priors flexibles ; convivial R	Nécessite lat/lon/temps ; moins mature pour multivarié	Élevé (efficace TMB)	sdmTMB

TMB: Template Model Builder R package permettant d'ajuster efficacement des modèles complexes.

SPDE: Stochastic Partial Differential Equation, cadre mathématique utilisé pour construire et représenter des champs aléatoires spatiaux et spatio-temporels continus de manière efficace computationnel.

Aperçu des Méthodes Statistiques - plus détaillé

Aspect	GLM	GAM	GLMM	VAST	sdmTMB
Idée principale	Régression paramétrique de la CPUE (log-CPUE) sur des covariables avec distribution d'erreur et lien choisis.	Régression semi-paramétrique avec fonctions lissées pour des effets non linéaires des covariables.	GLM avec effets aléatoires (par ex. navire, zone)	Modèle delta-GLMM spatio-temporel basé sur TMB, avec champs aléatoires spatiaux et spatio-temporels.	GLMM SPDE basé sur TMB avec interface par formule et outils dédiés aux ind.
Usage	Standardisation « classique » de CPUE palangrière ou chalutière avec facteurs (année, zone, engin, navire).	Standardisation de CPUE quand les réponses aux covariables / saisons sont fortement non linéaires.	Standardisation de CPUE incluant effets aléatoires navire, marée ou zone.	Standardisation spatio-temporelle de CPUE ou d'ind de campagne, y compris plusieurs flottilles/espèces.	Standardisation spatio-temporelle de CPUE / ind de océanographique campagne.
Structure spatiale	Uniquement via des facteurs de zone ou coordonnées simples ; pas de corrélation spatiale explicite.	Peut inclure des lissages spatiaux, mais pas de champ géostatistique complet.	Peut inclure des effets aléatoires de zone ; toujours sans champ géostatistique explicite.	Champs aléatoires spatiaux et spatio-temporels explicites via SPDE ; modèle delta pour rencontre / CPUE positive.	Champs SPDE spatiaux et spatio-temporels (comme VAST) avec familles delta ou Tweedie.
Traitement de la surabondance de zéros	Via la famille choisie (par ex. delta-lognormale, Poisson, NB) mais pas automatique.	Comme le GLM, éventuellement en delta ; flexible pour la non-linéarité mais pas spécialisé pour les zéros.	Comme le GLM, avec effets aléatoires	Modèle delta-GLMM natif (rencontre + composante positive) avec champs spatiaux pour chaque partie.	Support intégré pour modèles delta, Tweedie et autres familles adaptées aux CPUE à nombreux zéros.
Dynamique temporelle / spatio-temporelle	Année comme facteur ou covariable ; pas de corrélation temporelle explicite.	Année comme lissage ou facteur ; pas de corrélation temporelle explicite.	Effets aléatoires d'année possibles ; structure temporelle limitée.	Champ aléatoire spatio-temporel structuré (par ex. marche aléatoire ou AR1 entre années) plus effets d'année.	Comme VAST : champs aléatoires spatio-temporels plus effets d'année, définis dans la formule du modèle.
Indice standardisé	Coefficients d'année ou moyennes marginales utilisés comme indice de CPUE standardisée.	Effets partiels ou prédictions par année (souvent re-scalées) comme indice.	Effets d'année prédits (fixes + aléatoires) comme indice standardisé.	Indice annuel d'abondance/biomasse issu des prédictions spatiales et de l'intégration, plus centre de gravité et aire occupée.	fonction <code>get_index()</code> fournit l'indice annuel de biomasse/abondance avec IC.
Complexité	Simple, rapide, très répandu ; diagnostics faciles.	Complexité modérée ; choix des lissages, risque de sur-ajustement	Modéré ; estimation des effets aléatoires et convergence parfois délicates.	Plus complexe ; nécessite maille/nœuds et paramétrage mais efficace sur grands jeux de données via TMB.	Plus complexe que GLM/GAM mais interface assez conviviale ;
Avantages	Simplicité, très accepté en évaluation de stocks	Capture bien les effets non linéaires des covariables, meilleur ajustement que GLM pour la CPUE.	Prend en compte l'hétérogénéité navire ou zone ; plus réaliste qu'un GLM purement à effets fixes.	Meilleure prise en compte de l'autocorrélation spatiale et spatio-temporelle	Combine géostatistique type VAST avec une interface R friendly.
Limites	Pas de corrélation spatiale ; difficultés avec forte non-linéarité et zéro-inflation complexe.	Pas de champ géostatistique explicite, interprétation moins simple avec de nombreux lissages.	Corrélation spatiale seulement approximée via effets aléatoires de zone ; moins puissant que les modèles géostatistiques.	Courbe d'apprentissage plus raide ; plus de paramètres	Récemment, Emergent

Aperçu des Méthodes Statistiques



- **GLM** (Modèles Linéaires Généralisés) - les plus courants

```
glm(log_cpue + 1 ~ year + quarter + area + vessel + gear + bait,  
     family = gaussian(link = "identity"), data = cpue_data)
```

[Hashimoto et al., 2025](#)

- **GAM** (Modèles Additifs Généralisés) - relations non linéaires

```
gam(log_cpue + 1 ~ year + s(lon, lat, bs = "tp") + s(month, bs = "cc") +  
     s(depth) + area + vessel,  
     family = gaussian, data = cpue_data)
```

[Itoh and Takahashi,
2025](#)

- **GLMM** (Modèles Linéaires Généralisés Mixtes) - effets aléatoires

```
glmer(log_cpue ~ year + quarter + area + (1|vessel) + (1|area),  
       family = gaussian, data = cpue_data)
```

[Correa et al., 2024](#)

- **Spatial autocorrelation**

R library VAST (Vector Autoregressive Spatio-Temporal)

R library sdmTMB (species distribution model based Template Model Builder)

Aperçu des Méthodes Statistiques



Autocorrelation spatial: [Che et al., 2025](#)

library VAST

```
fit_vast <- VAST(data_list = VAST_data, family = c("lognormal", "gamma"), fieldConfig = c("Index" =  
c(Index_cpt = 2)), X1_formula = ~ 0 + sst, X2_formula = ~ 0 + sst + depth, vessel_effects = TRUE, Lon_i =  
VAST_data$Lon_i, Lat_i = VAST_data$Lat_i, getReportCovariance = FALSE)
```

library sdmTMB

```
sdmTMB(cpue_k.h ~ 0 + (1 | fvessel_ID) + s(log.depth_scaled, k=3) + as.factor(year), mesh = mesh.F,  
time = "year", spatial = "on", spatiotemporal = "iid", tweedie(link = "log"), data = data)
```

Aperçu des Méthodes Statistiques



Les modèles base (GLM/GAM/GLMM) fonctionnent pour les données non spatiales et les modèles avancés (sdmTMB/VAST) corrigent l'autocorrélation spatiale.

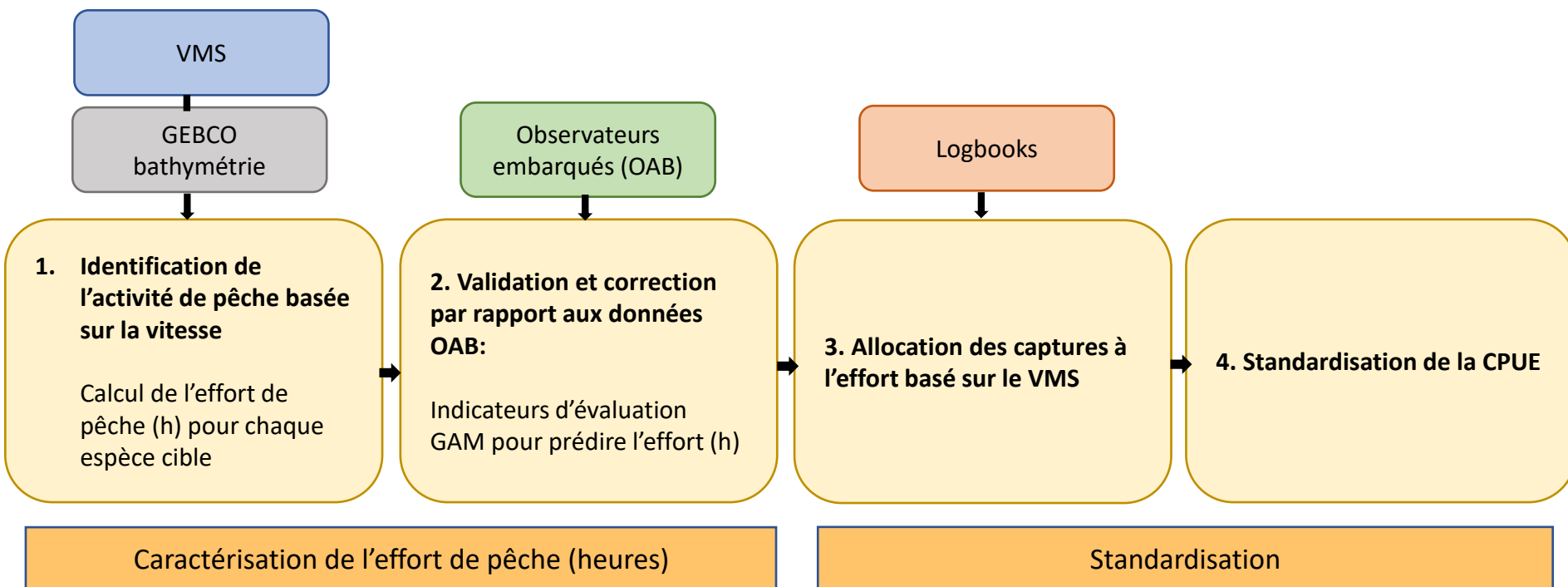
VAST étant plus performant pour les réponses multiples et sdmTMB plus simple R syntaxe/rapide pour les variables univariées.

Le choix dépend de l'échelle des données, de l'étendue spatiale et des besoins.

Messages clés à retenir

- La standardisation des CPUE est essentielle pour des indices d'abondance fiables.
- Les méthodes appropriées dépendent de la disponibilité des données et des caractéristiques de la pêcherie.
- Critique pour l'évaluation des stocks à données limitées.

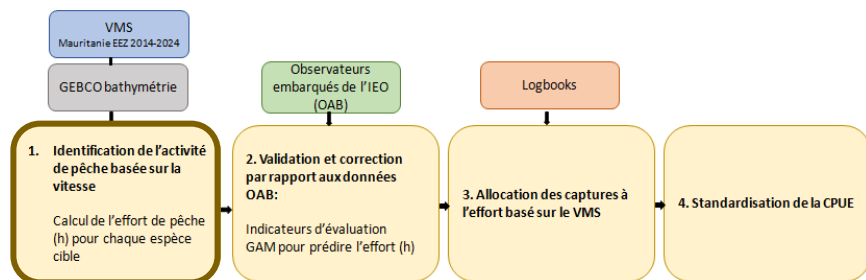
3 Take Home
Messages



Études de cas : Mauritanie, Guinée-Bissau, Sénégal, Guinée, Angola

Étude de cas: MAURITANIE

- Flotte : Chalutiers crevettiers espagnols dans la ZEE mauritanienne (opérant dans le cadre de l'APPD)
- Espèces cibles multiples: jusqu'à trois espèces cibles exploitées dans différentes gammes de profondeur au cours d'une même journée de pêche
- Données VMS pour l'ensemble de la flotte (espagnole)
- Données des OAB de l'IEO
- Données de journaux de bord agrégées : uniquement informations journalières de capture
- CPUE par espèce actuellement exprimée en kg/jour de pêche



1. Identification de l'activité de pêche basée sur la vitesse

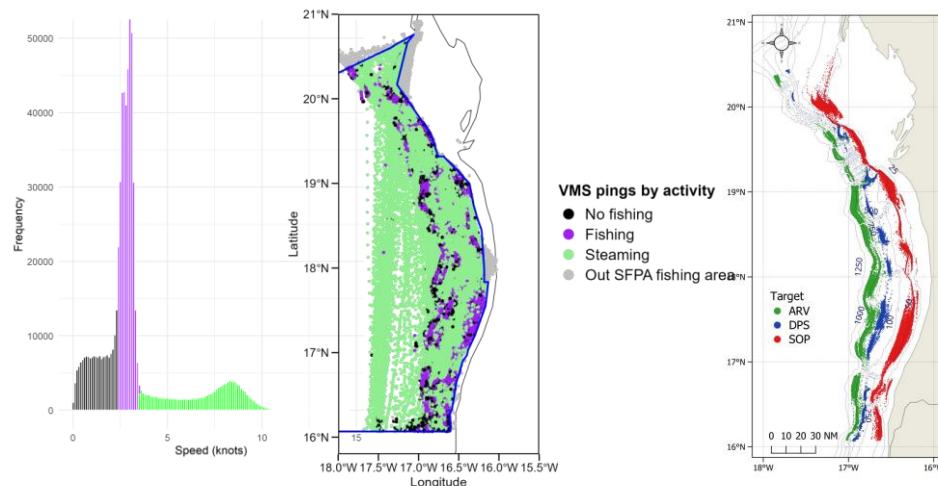
Plage de vitesse de pêche : 2.36–3.49 nœuds

Filtrer les positions en statut « pêche » et attribuer à chaque espèce cible selon la gamme de profondeur

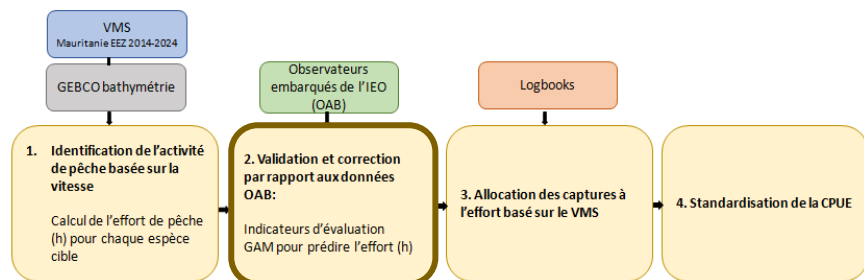
Penaeus notialis - SOP: 10-80m

Parapenaeus longirostris – DPS : 150-350m

Aristeus varidens – ARV : 400-1000m



WP 4: Caractérisation de l'effort de pêche et standardisation de la CPUE

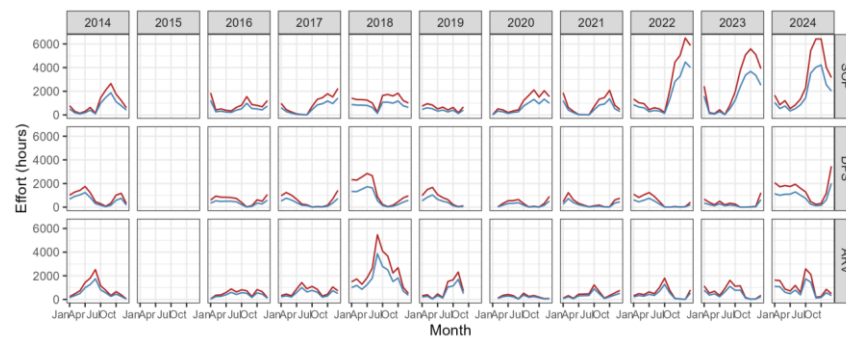
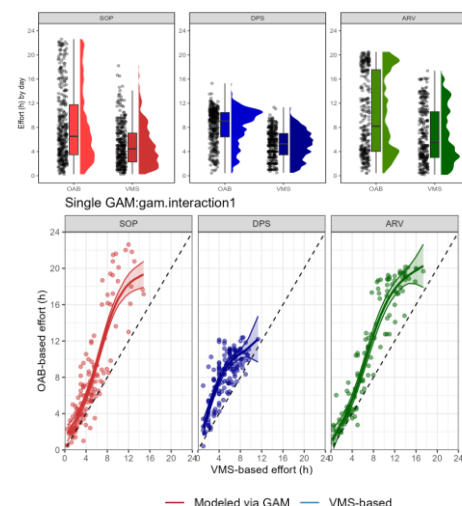


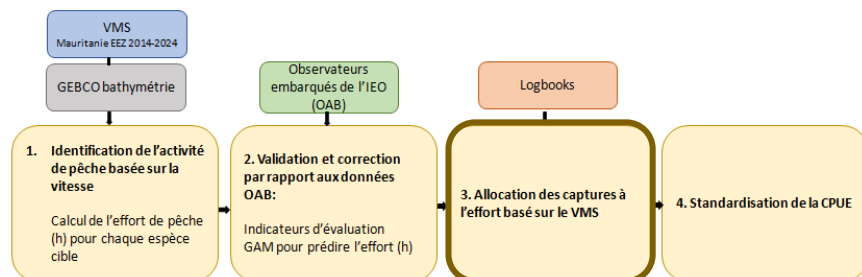
2. Validation et correction par rapport aux données OAB

Indicateurs d'évaluation : Exactitude, Précision, Rappel, F1Score

Modèle GAM pour « corriger » et obtenir l'effort prédit

Calcul de l'effort (heures)



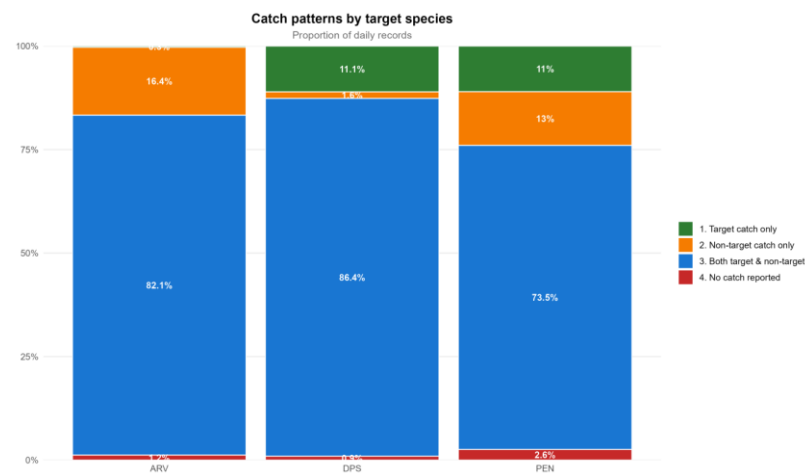


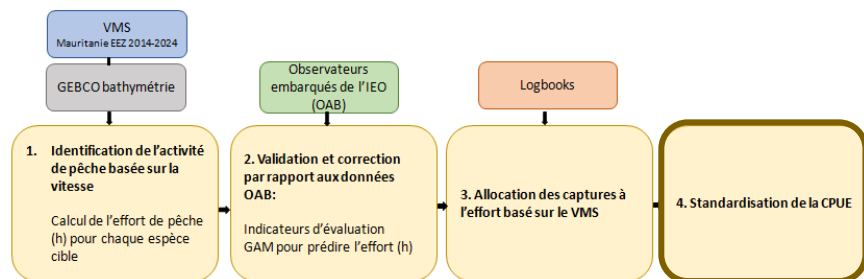
3. Allocation des captures à l'effort basé sur le VMS

Effort basé sur le VMS calculé par espèce cible, jour et navire

L'effort est agrégé et les coordonnées du centroïde sont conservées

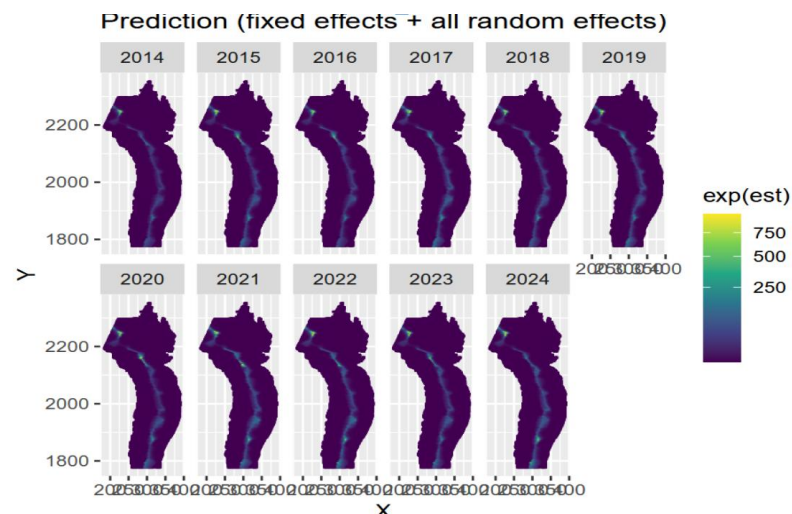
Capture journalière allouée par espèce cible





4. Standardisation de la CPUE pour chaque stock cible

Modélisation spatio-temporelle sdmTMB



Étude de cas: GUINÉE-BISSAU

- Flotte : Chalutiers crevettiers espagnols dans la ZEE de Guinée-Bissau (opérant dans le cadre de l'APPD)
- Espèces cibles multiples : jusqu'à trois espèces cibles exploitées dans différentes gammes de profondeur au cours d'une même journée de pêche
- Données VMS pour l'ensemble de la flotte (espagnole)
- Données des OAB de l'IEO + PAMBAS
- Données de journaux de bord agrégées : uniquement informations journalières de capture
- CPUE par espèce actuellement exprimée en kg/jour de pêche

Étude de cas: **SÉNÉGAL**

- Flotte : Chalutiers crevettiers dans la ZEE de Sénégal
- Espèces cibles multiples : jusqu'à deux espèces cibles exploitées dans différentes gammes de profondeur au cours d'une même journée de pêche
- Données VMS pour l'ensemble de la flotte??
- Données des OAB du CRODT - **PAMBAS**
- **Données détaillées de journaux de bord par opération de pêche**
- CPUE gamba actuellement standardisée avec un **delta-GLM** tester sdmTMB?

Étude de cas: GUINÉE

- Flotte : Chalutiers crevettiers espagnols dans la ZEE de Guinée (opérant sous autorisation directe certaines années)
+ autres???
- Espèces cibles multiples : jusqu'à deux espèces cibles exploitées dans différentes gammes de profondeur au cours d'une même journée de pêche
- Données VMS – uniquement pour les navires espagnols??
- Données des OAB de l'IEO – uniquement pour les navires espagnols et PAMBAS
- Données de journaux de bord agrégées – uniquement pour les navires espagnols??
- CPUE par espèce actuellement exprimée en kg/jour de pêche

Étude de cas: **ANGOLA**

- Flotte : Chalutiers crevettiers espagnols dans la ZEE de l'Angola (opérant sous autorisation directe) + flotte angolaise + autres??
- Espèces cibles multiples : jusqu'à deux espèces cibles exploitées dans différentes gammes de profondeur au cours d'une même journée de pêche
- Données VMS – uniquement pour les navires espagnols??
- Données des OAB de l'IEO – uniquement pour les navires espagnols et PAMBAS
- Données de journaux de bord agrégées – uniquement pour les navires espagnols??
- CPUE par espèce actuellement exprimée en kg/jour de pêche