



Modélisation spatio-temporelle des zones fonctionnelles halieutiques par combinaison des données de captures à échelle fine des navires de pêche et des navires de recherche halieutique

Baptiste Alglave (Doctorant)

Encadrement : Etienne Rivot, Youen Vermard,
Mathieu Woillez, Marie-Pierre Etienne

Ifremer Nantes – Unité EMH
Agrocampus Ouest

Contexte écologique et réglementaire

La bonne gestion des ressources halieutiques

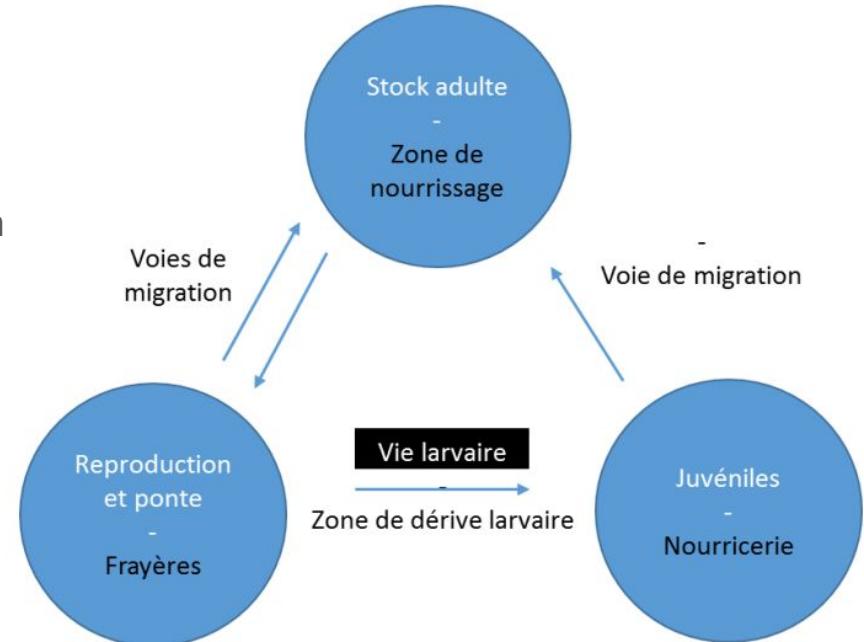
→ Connaissances des zones clés pour la réalisation du cycle de vie des espèces

- Frayères
- Nourriceries
- Voie de migration

⇒ **Zones Fonctionnelles Halieutiques d'importance (ZFI)**



Identification et caractérisation de ces zones dans le temps et l'espace ?



Représentation schématique du cycle de vie d'une espèce d'intérêt halieutique.
(Harden Jones, 1968 in Delage and Le Pape, 2016)

Les données à disposition

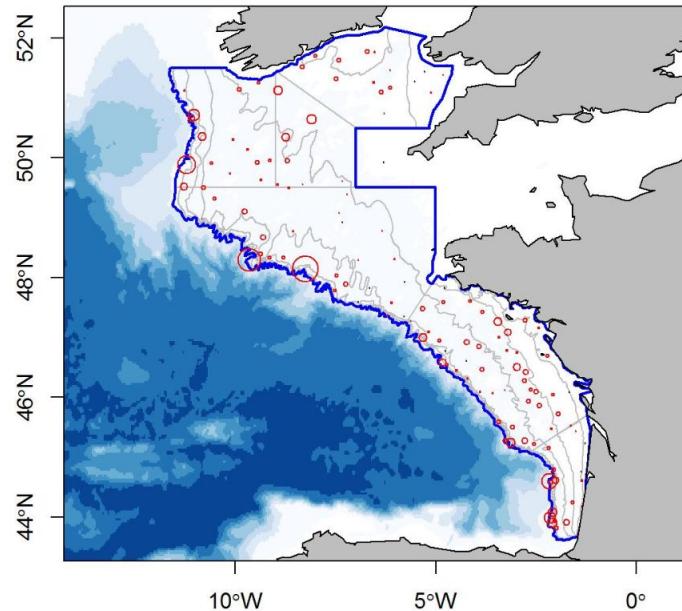
Données
scientifiques



Schéma d'échantillonnage et
protocoles standardisés



Coûteux
Couverture spatio-temporelle limitée



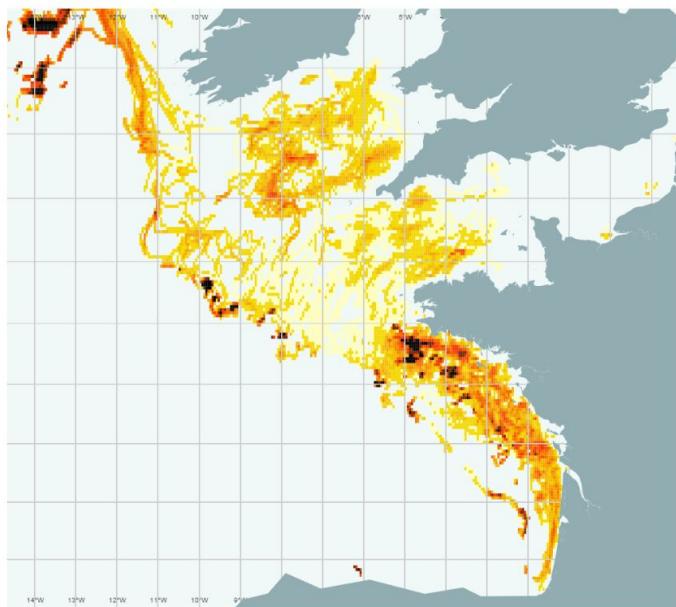
Les données à disposition

Données logbooks
x
Données VMS



Données disponibles sur toute l'année
Peu coûteux

Hypothèses pour répartir les captures sur les
points VMS
Données biaisées par le comportement
de pêche



Distribution spatiale du merlu
Données VMS (oct. 2012)
Woillez (com.perso.)

Contexte

Sujet de
thèse

Modèle et
hypothèse

Échantillonnage
préférentiel

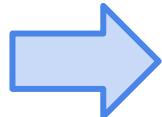
Simulations

Bilan

Sujet de thèse :

Modélisation spatio-temporelle des zones fonctionnelles halieutiques par combinaison des données de captures à échelle fine des navires de pêche et des navires de recherche halieutique.

Problématique :



Peut-on utiliser la donnée commerciale pour inférer sur la dynamique spatio-temporelle des espèces d'intérêts halieutiques et pour identifier des zones fonctionnelles halieutiques d'importance ?

Une thèse à 4 composantes :

Méthodologique

Écologique

Halieutique

Gestion et
conservation

Contexte

Sujet de
thèse

Modèle et
hypothèses

Échantillonnage
préférentiel

Simulations

Bilan

Le modèle pour répondre à la problématique

Hyper-paramètres

Variables latentes

Observations

Contexte

Sujet de
thèse

Modèle et
hypothèses

Échantillonnage
préférentiel

Simulations

Bilan

Le modèle pour répondre à la problématique

Hyper-paramètres

Variables latentes

$d(t,x,a)$

Variable latente

Densité d'individus :

- à un temps t
- à un point x
- pour une classe d'âge a

Observations

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

Le modèle pour répondre à la problématique

Hyper-paramètres

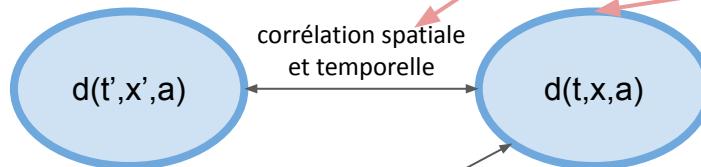
Corrélations

$$\rho(t,x,a)$$

Erreurs de processus

$$\sigma^2$$

Variables latentes



Hypothèses sur la structure du champs latent

- Corrélation spatio-temporelle
- Variables explicatives (covariables environnementales)

Observations

Covariab...
(environnementales ou autres)

Contexte

Sujet de
thèse

Modèle et
hypothèses

Échantillonnage
préférentiel

Simulations

Bilan

Le modèle pour répondre à la problématique

Hyper-paramètres

Erreurs d'observations

λ

Corrélations

$\rho(t,x,a)$

Erreurs de processus

σ^2

Variables latentes

$d(t',x',a)$

$d(t,x,a)$

corrélation spatiale
et temporelle

Observations

Covariables
(environnementales ou
autres)

Campagnes océanographiques

Observations
Données de campagnes

Kristensen et al. (2014)

Le modèle pour répondre à la problématique

Hyper-paramètres

Erreurs d'observations

λ

Corrélations

$\rho(t,x,a)$

Erreurs de processus

σ^2

Variables latentes

$d(t',x',a)$

$d(t,x,a)$

Observations
Données commerciales

corrélation spatiale
et temporelle

capturabilité

Modèles de comportement de pêche

Observations

Covariables
(environnementales ou autres)

$C_{\text{campagnes}}(t,x,a)$

$C_{\text{VMS}}(t,x,a)$

Campagnes océanographiques

Données VMS x
logbooks

Kristensen et al. (2014)
Conn et al. (2017)
Pennino et al. (2018)

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

Le modèle pour répondre à la problématique

Hyper-paramètres

Erreurs d'observations

Corrélations

Erreurs de processus

 λ $\sigma(t \mid a)$ σ^2

Variables latentes

 $d(t', x, a)$

Hypothèse

Répartition préférentielle des coups de pêche

Observations

capturabilité

Covariables
(environnementales ou autres)

 $C_{\text{campagnes}}(t, x, a)$

Modèles de comportement de pêche

 $C_{\text{VMS}}(t, x, a)$

Données VMS x logbooks

Campagnes océanographiques

Kristensen et al. (2014)
Conn et al. (2017)
Pennino et al. (2018)

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

Densité d'individus Nb de coups de pêche en 1 point Captures

Probabilité conjointe :

- de la variable latente **S**
- du processus d'échantillonnage **X**
- des observations **Y**

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

Modèle :

Hyp 1 : S est un processus gaussien stationnaire de moyenne 0, de variance σ^2 de fonction de corrélation $\rho(u, \phi) = \text{corr}\{S(x), S(x')\}$ pour tout x et x' distants de u .

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↑
Densité Nb de coups Captures
d'individus de pêche en 1 point

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

Modèle :

Hyp 1 : S est un processus gaussien stationnaire de moyenne 0, de variance σ^2 de fonction de corrélation $p(u, \phi) = \text{corr}\{S(x), S(x')\}$ pour tout x et x' distants de u .

Hyp 2 : Conditionnellement sur S , X est un processus de Poisson non-homogène d'intensité :
 $\lambda(x) = \exp\{\alpha + \beta S(x)\}$

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↑
Densité Nb de coups Captures
d'individus de pêche en 1 point

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

Modèle :

Hyp 1 : S est un processus gaussien stationnaire de moyenne 0, de variance σ^2 de fonction de corrélation $p(u, \phi) = \text{corr}\{S(x), S(x')\}$ pour tout x et x' distants de u .

Hyp 2 : Conditionnellement sur S , X est un processus de Poisson non-homogène d'intensité : $\lambda(x) = \exp\{\alpha + \beta S(x)\}$

Hyp 3 : Conditionnellement sur S et X , Y est un ensemble de variables gaussiennes mutuellement indépendantes t.q. $Y | xi \sim N(\mu + S(xi), \tau^2)$

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↑
 Densité Nb de coups Captures
 d'individus de pêche en 1 point

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↑
 Densité Nb de coups Captures
 d'individus de pêche en 1 point

Modèle :

Hyp 1 : S est un processus gaussien stationnaire de moyenne 0, de variance σ^2 de fonction de corrélation $p(u, \phi) = \text{corr}\{S(x), S(x')\}$ pour tout x et x' distants de u .

Hyp 2 : Conditionnellement sur S , X est un processus de Poisson non-homogène d'intensité : $\lambda(x) = \exp\{\alpha + \beta S(x)\}$

Hyp 3 : Conditionnellement sur S et X , Y est un ensemble de variables gaussiennes mutuellement indépendantes t.q. $Y_i | x_i \sim N(\mu + S(x_i), \tau^2)$



Généralisé et repris par Pati et al. (2011), Conn et al. (2017)

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$$P[S, X, Y]$$

Variable latente
Points échantillonés Observations

Modèle de base

Equation de processus

Densité

covariables

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

effet aléatoire spatial

Modèles d'observations des données commerciales

Modèles d'observations des données scientifiques

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$P[S, X, Y]$

Variable latente
Points échantillonnés Observations

Modèle de base

Equation de processus

Densité

covariables

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

effet aléatoire spatial

Modèles d'observations des données commerciales

Processus d'échantillonnage de la donnée scientifique

Indépendant du champs latent

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$P[S, X, Y]$

Variable latente
Points échantillonnés Observations

Modèle de base

Equation de processus

Densité

covariables

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

effet aléatoire spatial

Modèles d'observations des données commerciales

Processus d'échantillonnage de la donnée scientifique

Indépendant du champs latent

Équation d'observation de la donnée scientifique

Modèle d'observation standard

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$$P[S, X, Y]$$

Variable latente
Points échantillonnés Observations

Modèle de base

Equation de processus

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

Densité

covariables

effet aléatoire spatial

Processus d'échantillonnage de la donnée commerciale

$$X_{com} \sim \mathcal{P}(\lambda(x))$$

Processus d'échantillonnage de la donnée scientifique

Indépendant du champs latent

Équation d'observation de la donnée scientifique

Modèle d'observation standard

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$$P[S, X, Y]$$

Variable latente
Points échantillonnés Observations

Modèle de base

Equation de processus

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

Processus d'échantillonnage de la donnée commerciale

$$\lambda(x) = \exp \{ \alpha_X + b \cdot S(x) + \gamma_X(x)^T \cdot \beta_X + \eta(x) \}$$

$$X_{com} \sim \mathcal{P}(\lambda(x))$$

Densité

covariables

effet aléatoire spatial

Processus d'échantillonnage de la donnée scientifique

Indépendant du champs latent

Équation d'observation de la donnée scientifique

Modèle d'observation standard

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$$P[S, X, Y]$$

Variable latente
Points échantillonnés
Observations

Modèle de base

Equation de processus

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

Processus d'échantillonnage de la donnée commerciale

$$\lambda(x) = \exp \{ \alpha_X + b \cdot S(x) + \gamma_X(x)^T \cdot \beta_X + \eta(x) \}$$

$$X_{com} \sim \mathcal{P}(\lambda(x))$$

Équation d'observation de la donnée commerciale

Modèle d'observation standard

Densité

covariables

effet aléatoire spatial

Processus d'échantillonnage de la donnée scientifique

Indépendant du champs latent

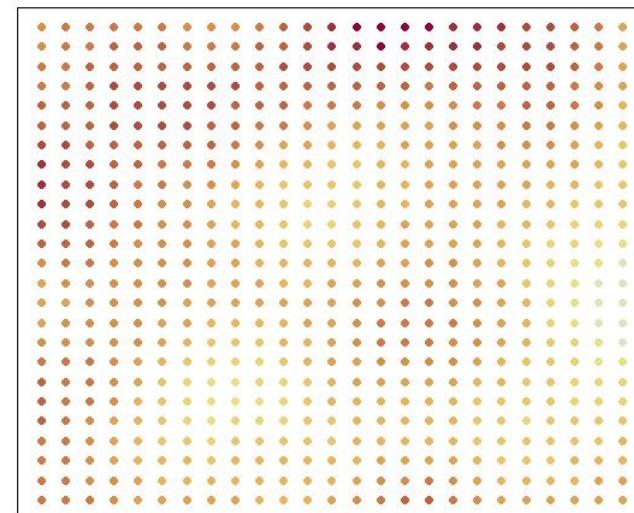
Équation d'observation de la donnée scientifique

Modèle d'observation standard

Descriptions des simulations

Configurations des simulations

Basé sur les codes de Conn et al. (2017)



Simulation d'un champs aléatoire
dans une grille de 625 cellules

- Domaine discret : grille de 25 cellules sur 25 cellules (625 cellules)
- Simulation du champs latent, des observations et des covariables :
⇒ Package RandomField
 - 50 observations scientifiques
 - 1500 observations commerciales
 - 3 niveaux d'échantillonnage préférentiel (b) : 0 (faible), 1 (modéré), 3 (élevé)

Contexte

Sujet de
thèse

Modèle et
hypothèses

Échantillonnage
préférentiel

Simulations

Bilan

Descriptions des simulations

Basé sur les codes de Conn et al. (2017)

Modèles d'estimations :

⇒ Package TMB

- Modèle s'ajustant uniquement à la **donnée commerciale**
- Modèle s'ajustant uniquement à la **donnée scientifique**
- Modèle **combinant les deux sources de données**

Descriptions des simulations

Basé sur les codes de Conn et al. (2017)

Modèles d'estimations :

⇒ Package TMB

- Modèle s'ajustant uniquement à la **donnée commerciale**
- Modèle s'ajustant uniquement à la **donnée scientifique**
- Modèle **combinant les deux sources de données**

Critères d'évaluation des différents modèles :

$$\text{Calcul de l'abondance totale} \quad Ab = \sum_{i=1}^n S_i$$

- Biais relatif de l'abondance totale

$$\text{Biais Relatif} = \frac{(Ab - \widehat{Ab})}{Ab}$$

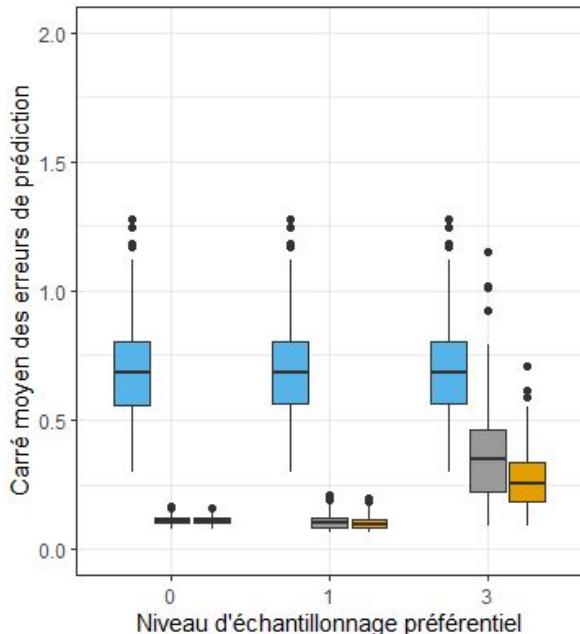
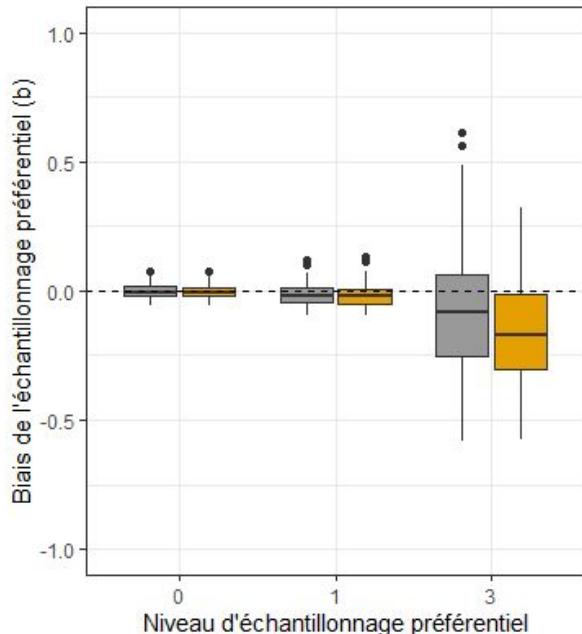
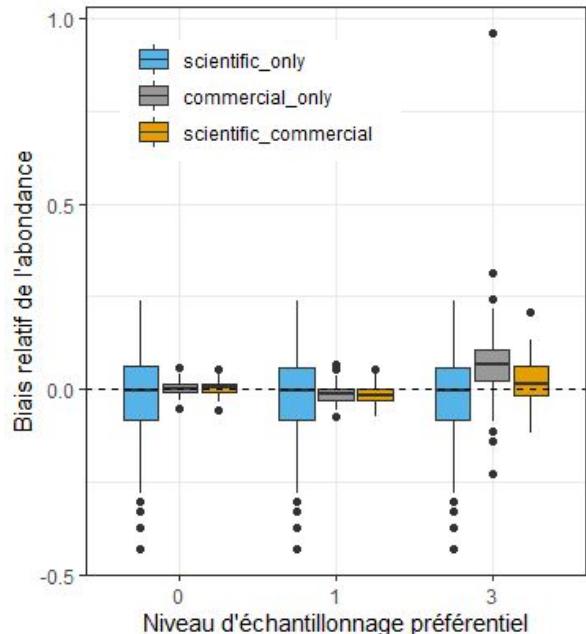
- Biais du paramètre associé à l'échantillonnage préférentiel (b)

$$\text{Biais de } b = b - \widehat{b}$$

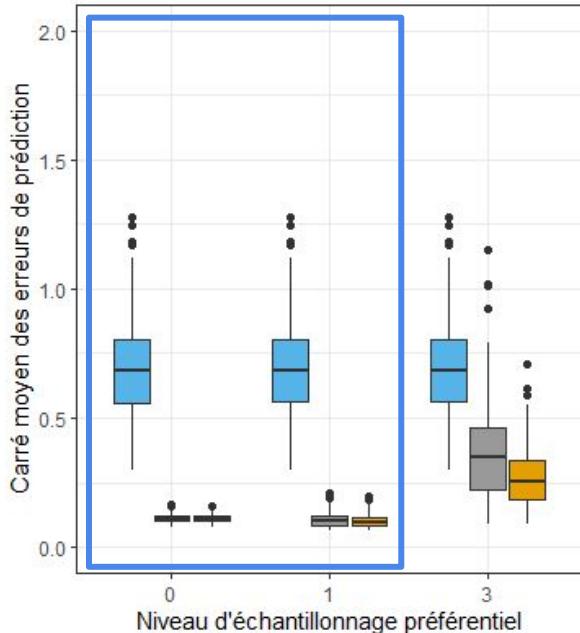
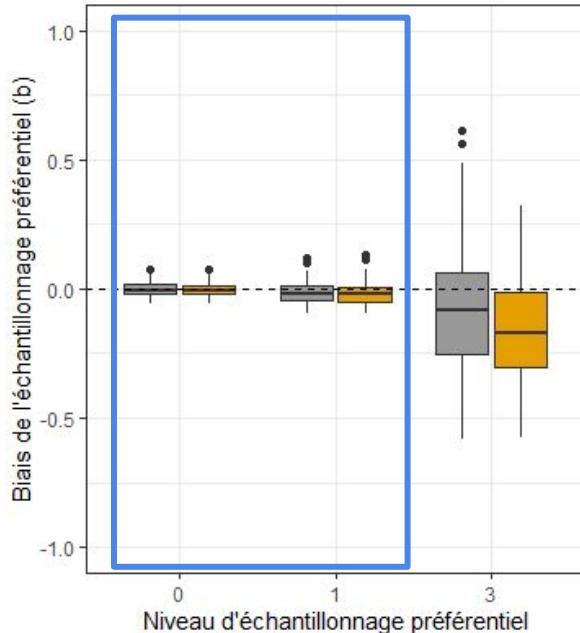
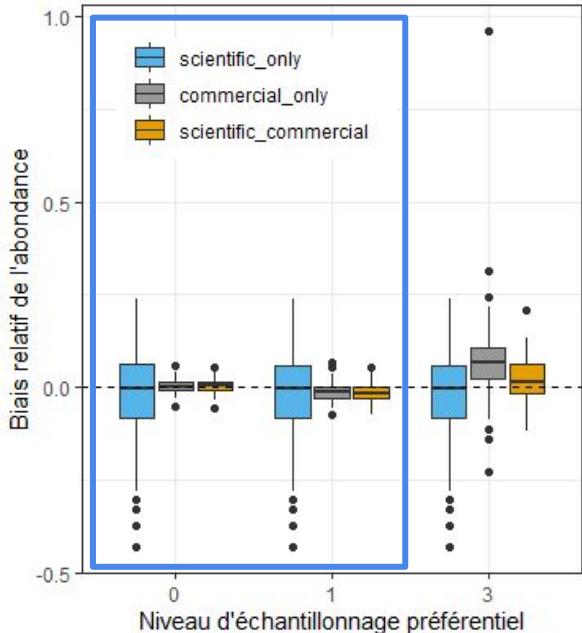
- Carré moyen des erreurs de prédiction

$$CMEP = \frac{1}{n} \sum_{i=1}^n (S_i - \widehat{S}_i)^2$$

Résultats de simulation

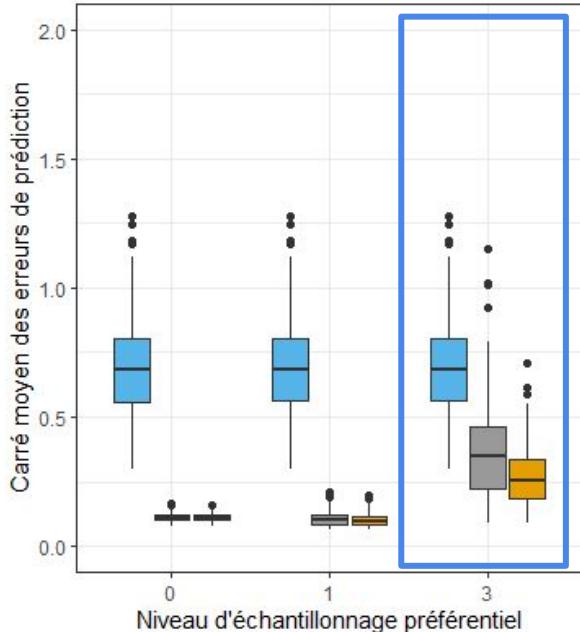
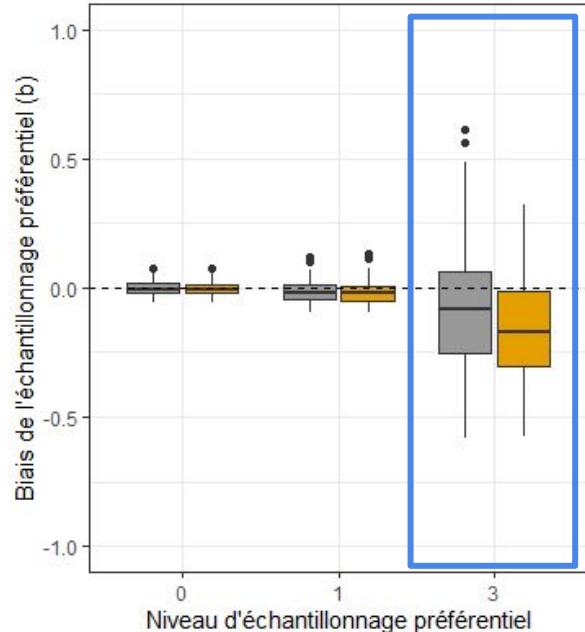
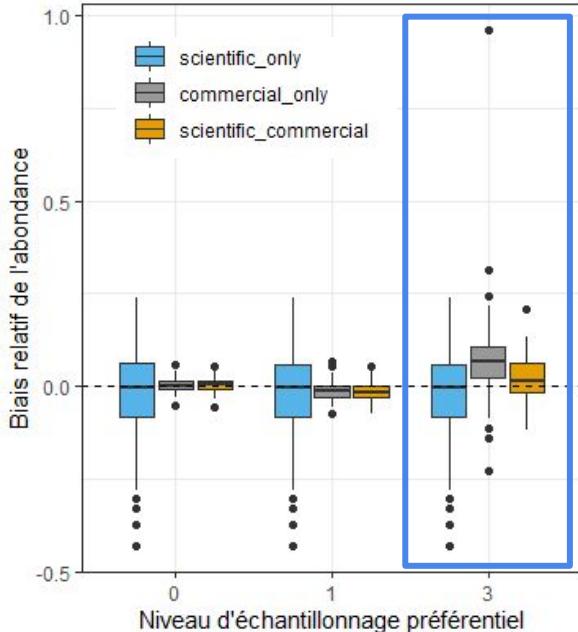


Résultats de simulation



Pour un niveau d'échantillonnage préférentiel nul ou modéré

Résultats de simulation



Pour un niveau d'échantillonnage préférentiel élevé

Discussions des résultats

Conclusion

- La donnée commerciale :
 - ⇒ source de donnée très intéressante dans le cas où l'échantillonnage n'est pas ou peu préférentiel
- L'approche intégrée :
 - ne corrige pas nécessairement le biais associé à l'échantillonnage préférentiel
 - peut fournir des prédictions plus précises

Discussions des résultats

Conclusion

- La donnée commerciale :
 - ⇒ source de donnée très intéressante dans le cas où l'échantillonnage n'est pas ou peu préférentiel
- L'approche intégrée :
 - ne corrige pas nécessairement le biais associé à l'échantillonnage préférentiel
 - peut fournir des prédictions plus précises

Perspectives

- Affiner les paramètres de simulations ⇒ analyse approfondie des deux jeux de données
- Quid des erreurs de déclaration et de la qualité des données commerciales ?
- Prochaine étape : ajuster aux données réelles.

Conclusions générale sur la thèse

- Innovation majeure ⇒ l'approche intégrée
 - Jusqu'à quel niveau de précision peut-on décrire la dynamique spatio-temporelle des espèces exploitées ?
 - Arrive-t-on à définir des ZFH d'importance ?
- De nombreux travaux ont déjà été réalisés
 - s'appuyer dessus !
 - beaucoup de perspectives !



Merci de votre attention

Bibliographie

DELAGE, Nicolas et LE PAPE, Olivier. Inventaire des zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française. Première partie: définitions, critères d'importance et méthode pour déterminer des zones d'importance à protéger en priorité. 2016.

DIGGLE, Peter J., MENEZES, Raquel, et SU, Ting-li. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 2010, vol. 59, no 2, p. 191-232.

CONN, Paul B., THORSON, James T., et JOHNSON, Devin S. Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. *Methods in Ecology and Evolution*, 2017, vol. 8, no 11, p. 1535-1546.

GERRITSEN, Hans D., MINTO, Caoilín, et LORDAN, Colm. How much of the seabed is impacted by mobile fishing gear? Absolute estimates from Vessel Monitoring System (VMS) point data. *ICES Journal of Marine Science*, 2013, vol. 70, no 3, p. 523-531.

GERRITSEN, Hans et LORDAN, Colm. Integrating vessel monitoring systems (VMS) data with daily catch data from logbooks to explore the spatial distribution of catch and effort at high resolution. *ICES Journal of Marine Science*, 2010, vol. 68, no 1, p. 245-252.

KRISTENSEN, Kasper, THYGESEN, Uffe Høgsbro, ANDERSEN, Ken Haste, et al. Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal of Fisheries and Aquatic Sciences*, 2013, vol. 71, no 2, p. 326-336.

MURRAY, Lee G., HINZ, Hilmar, HOLD, Natalie, et al. The effectiveness of using CPUE data derived from Vessel Monitoring Systems and fisheries logbooks to estimate scallop biomass. *ICES Journal of Marine Science*, 2013, vol. 70, no 7, p. 1330-1340.

PATI, Debdeep, REICH, Brian J., et DUNSON, David B. Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, 2011, vol. 98, no 1, p. 35-48.

PENNINO, Maria Grazia, PARADINAS, Iosu, ILLIAN, Janine B., et al. Accounting for preferential sampling in species distribution models. *Ecology and evolution*, 2019, vol. 9, no 1, p. 653-663.

REGIMBART, Amélie, GUITTON, Jérôme, et LE PAPE, Olivier. Zones fonctionnelles pour les ressources halieutiques dans les eaux sous souveraineté française. 2018.

THORSON, James T., FONNER, Robert, HALTUCH, Melissa A., et al. Accounting for spatiotemporal variation and fisher targeting when estimating abundance from multispecies fishery data. *Canadian Journal of Fisheries and Aquatic Sciences*, 2016, vol. 74, no 11, p. 1794-1807.

VERMARD, Youen, RIVOT, Etienne, MAHÉVAS, Stéphanie, et al. Identifying fishing trip behaviour and estimating fishing effort from VMS data using Bayesian Hidden Markov Models. *Ecological Modelling*, 2010, vol. 221, no 15, p. 1757-1769.

Annexes

Annexes I

Tout ce que je n'ai pas dit

Contexte

Sujet de
thèse

Modèle et
hypothèses

Échantillonnage
préférentiel

Simulations

Bilan

Les hypothèses

Les pêcheurs ciblent des zones (et des espèces) de façon préférentielle.

La donnée commerciale est informative de la dynamique spatio-temporelle des espèces d'intérêts halieutiques.

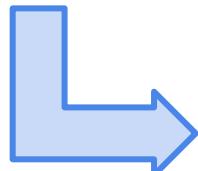
La donnée scientifique permet de corriger le biais introduit par l'échantillonnage préférentiel.

Les hypothèses

Les pêcheurs ciblent des zones (et des espèces) de façon préférentielle.

La donnée commerciale est informative de la dynamique spatio-temporelle des espèces d'intérêts halieutiques.

La donnée scientifique permet de corriger le biais introduit par l'échantillonnage préférentiel.



Si les hypothèses sont validées :

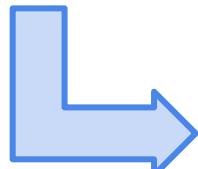
Prédictions précises et non-biaisées de la dynamique spatio-temporelle des espèces d'intérêts à une résolution plus fine qu'avec la donnée scientifique seulement.

Les hypothèses

Les pêcheurs ciblent des zones (et des espèces) de façon préférentielle.

La donnée commerciale est informative de la dynamique spatio-temporelle des espèces d'intérêts halieutiques.

La donnée scientifique permet de corriger le biais introduit par l'échantillonnage préférentiel.



Si les hypothèses sont validées :

Prédictions précises et non-biaisées de la dynamique spatio-temporelle des espèces d'intérêts à une résolution plus fine qu'avec la donnée scientifique seulement.

→ jusqu'à quel degré de précision peut-on décrire les ZFH ?

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

Densité d'individus
Nb de coups de pêche à un point donné
Captures

Probabilité conjointe de la variable latente, du processus d'échantillonnage et des observations

Cas de l'échantillonnage non-préférentiel (mais stochastique)

(X indép. de S , $P[S, X] = P[S] \cdot P[X]$)

$$P[S, X, Y] = P[S] \cdot P[X] \cdot P[Y|S(X)]$$

Processus d'échantillonnage

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↓
 Densité d'individus Nb de coups de pêche à un point donné Captures

Probabilité conjointe de la variable latente, du processus d'échantillonnage et des observations

Cas de l'échantillonnage déterministe

$$P[S, X, Y] = P[S] \cdot 1 \cdot P[Y|S(X)]$$

←
 Processus d'échantillonnage

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↓
 Densité d'individus Nb de coups de pêche à un point donné Captures

Probabilité conjointe de la variable latente, du processus d'échantillonnage et des observations

Cas de l'échantillonnage préférentiel
(X dépend de S)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

←
Processus d'échantillonnage

Approche théorique de l'échantillonnage préférentiel

L'approche de Diggle et al. (2010)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$

↑ ↑ ↓
Densité d'individus Nb de coups de pêche à un point donné Captures

Probabilité conjointe de la variable latente, du processus d'échantillonnage et des observations

Cas de l'échantillonnage déterministe

$$P[S, X, Y] = P[S] \cdot 1 \cdot P[Y|S(X)]$$

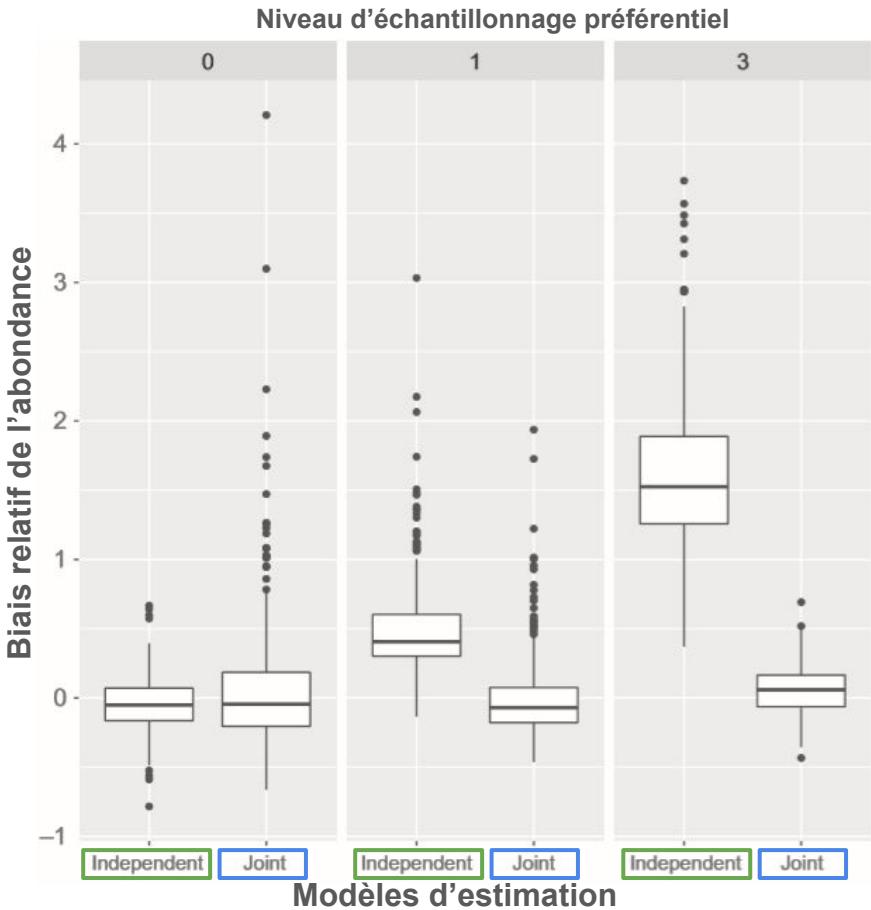
Cas de l'échantillonnage non-préférentiel (mais stochastique)
(X indép. de S, $P[X|S] = P[X]$)

$$P[S, X, Y] = P[S] \cdot P[X] \cdot P[Y|S(X)]$$

Processus d'échantillonnage

Cas de l'échantillonnage préférentiel
(X dépend de S)

$$P[S, X, Y] = P[S] \cdot P[X|S] \cdot P[Y|S(X)]$$



Performances des modèles intégrant l'échantillonnage préférentiel

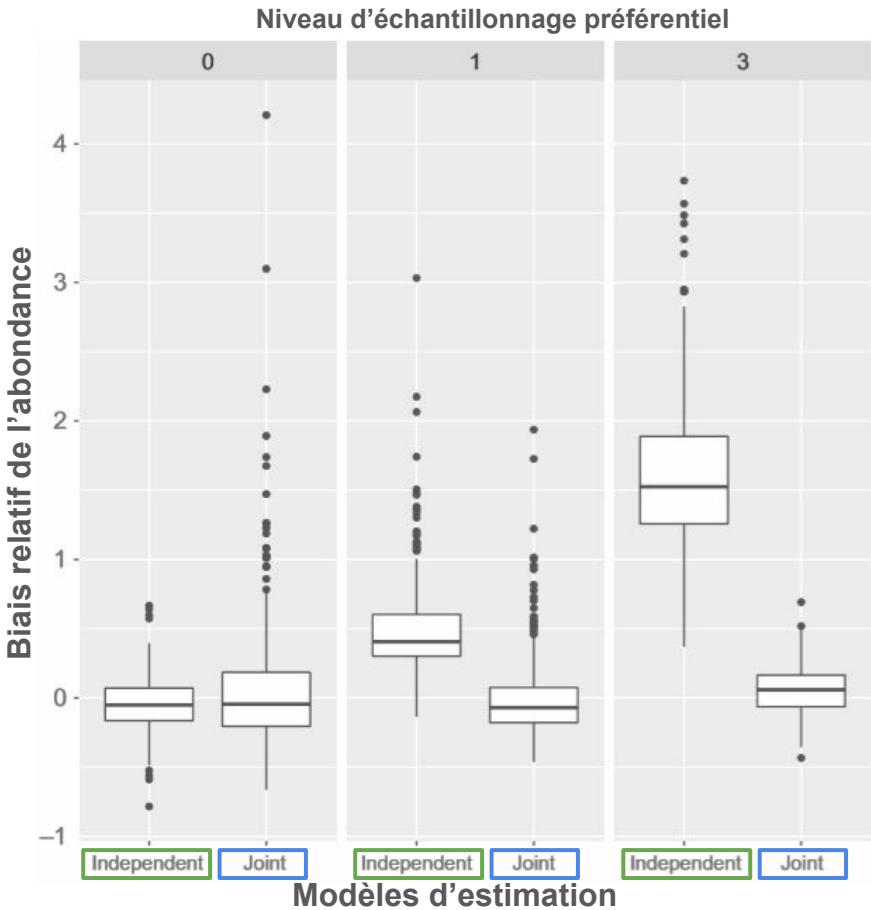
Les approches intégrant l'échantillonnage préférentiel réduisent le biais.

Approches instables.
Estimation pas toujours réaliste.

Erreur relative des prédictions d'abondances pour des données échantillonées de façon préférentielle pour différents niveaux d'échantillonnage préférentiel (Conn et al., 2017).

Modèles ne prenant pas en compte l'échantillonnage préférentiel

Modèles prenant en compte l'échantillonnage préférentiel



Performances des modèles intégrant l'échantillonnage préférentiel

Les approches intégrant l'échantillonnage préférentiel réduisent le biais.

Approches instables.
Estimation pas toujours réaliste.

Une approche intégrée permet-elle d'améliorer l'inférence sur les paramètres des modèles et sur l'abondance ?

Erreur relative des prédictions d'abondances pour des données échantillonées de façon préférentielle pour différents niveaux d'échantillonnage préférentiel (Conn et al., 2017).

Modèles ne prenant pas en compte l'échantillonnage préférentiel

Modèles prenant en compte l'échantillonnage préférentiel

Contexte

Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

$$P[S, X, Y]$$

Variable latente
Points échantillonnés
Observations

Modèle de base

Equation de processus

$$f(S(x)) = \alpha_S + \gamma_S(x)^T \cdot \beta_S + \delta(x)$$

Processus d'échantillonnage de la donnée commerciale

$$\lambda(x) = \exp \{ \alpha_X + b \cdot S(x) + \gamma_X(x)^T \cdot \beta_X + \eta(x) \}$$

$$X_{com} \sim \mathcal{P}(\lambda(x))$$

Équation d'observation de la donnée commerciale

$$Y_{com_i} | x_{com_i} \sim L_{com}(S(x_{com_i}), \sigma_1^2), i.i.d.$$

Loi de probabilité pour les modèles d'observations

effet aléatoire spatial

Processus d'échantillonnage de la donnée scientifique

Tirage aléatoire sans remise de p échantillons dans une grille à n cellules

Équation d'observation de la donnée scientifique

$$Y_{sci_i} | x_{sci_i} \sim L_{sci}(S(x_{sci_i}), \sigma_2^2), i.i.d.$$

Descriptions des simulations

Outils de simulation

- Simulation du processus latent, du processus d'échantillonnage et des covariables
→ Package RandomFields
- Ajustement du modèle
→ Package TMB/nlminb

Critères d'évaluation des différents modèles :

- Biais relatif de l'abondance
- Biais du paramètre associé à l'échantillonnage préférentiel (b)
- Carré moyen des erreurs de prédiction

Paramètres	Valeurs des paramètres
Dimension de la grille	625 cellules (25 x 25)
Modèles d'estimation	<ul style="list-style-type: none"> - Modèle s'ajustant uniquement à la donnée commerciale - Modèle s'ajustant uniquement à la donnée scientifique - Modèle combinant les deux sources de données
Nb d'observations pour la donnée scientifique	50
Nb d'observations pour la donnée commerciale	500 - 1000 - 1500
Échantillonnage préférentiel (b)	0 - 1 - 3
Nombre de simulations	100

Contexte

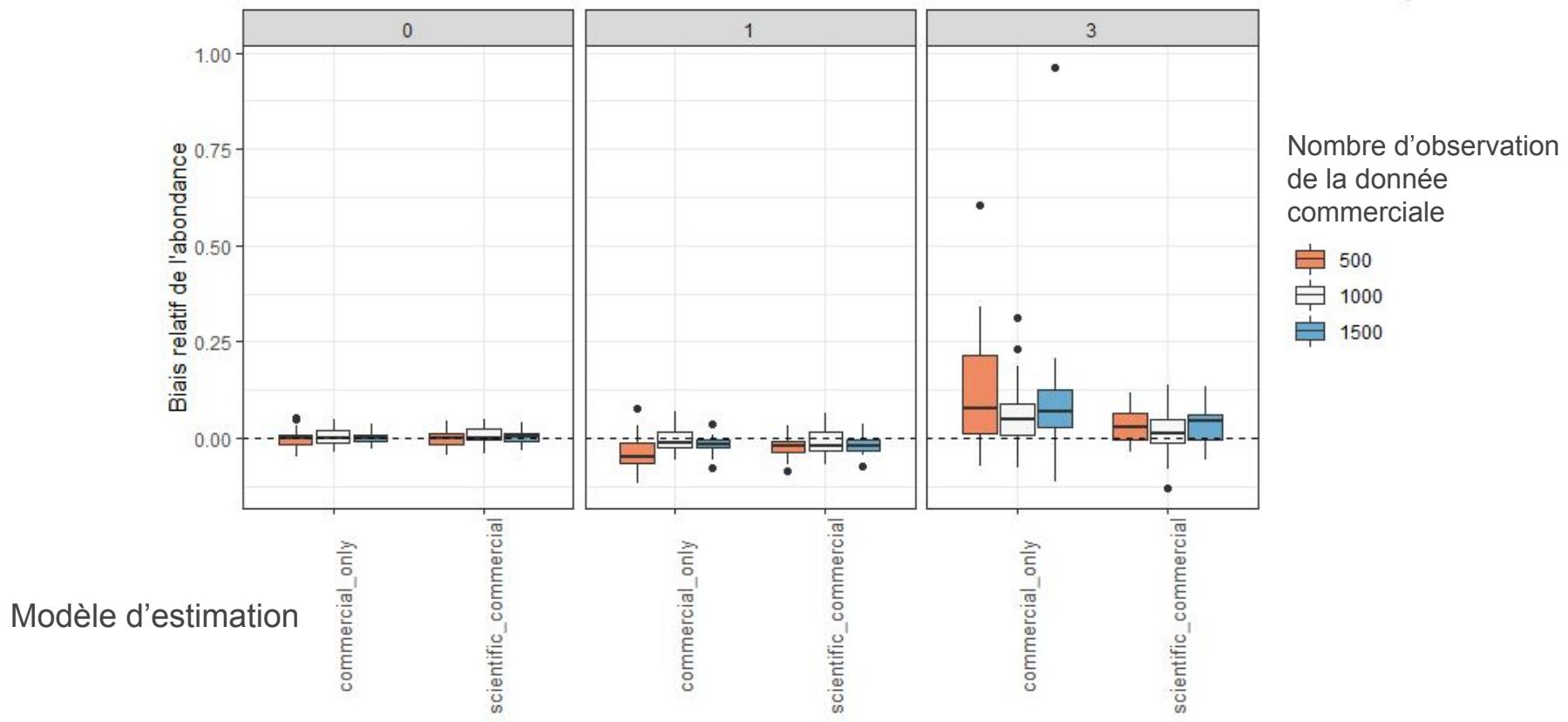
Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan



Contexte

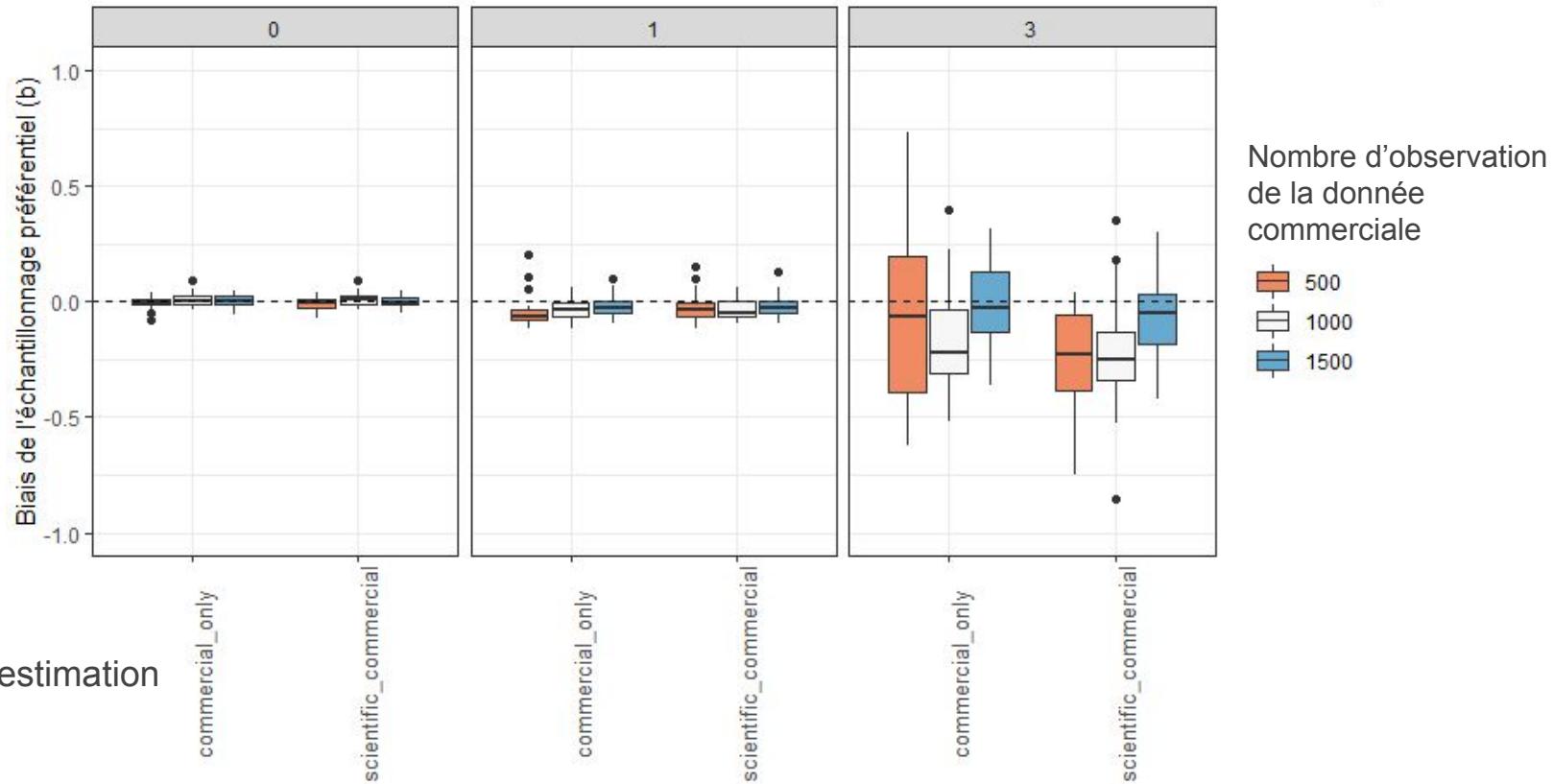
Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan



Contexte

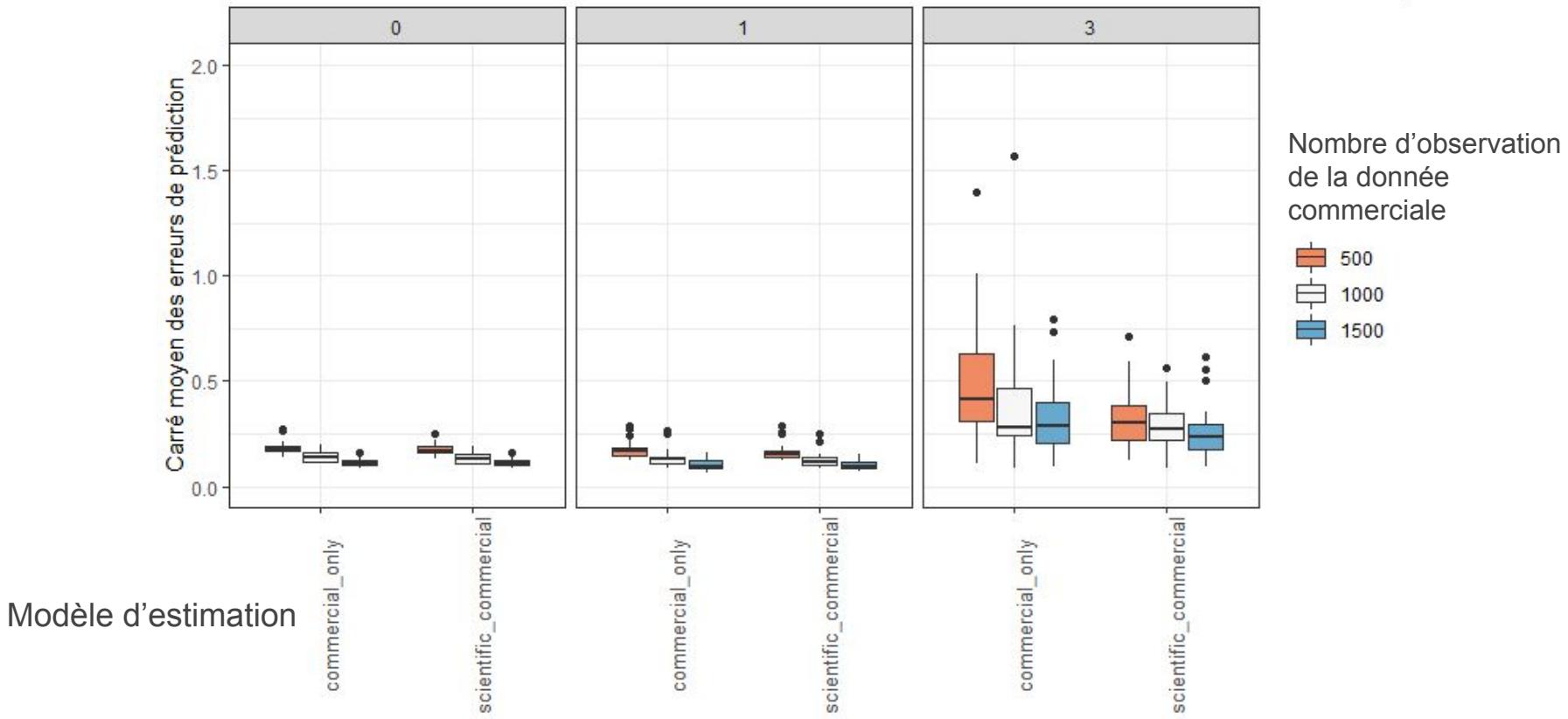
Sujet de thèse

Modèle et hypothèses

Échantillonnage préférentiel

Simulations

Bilan

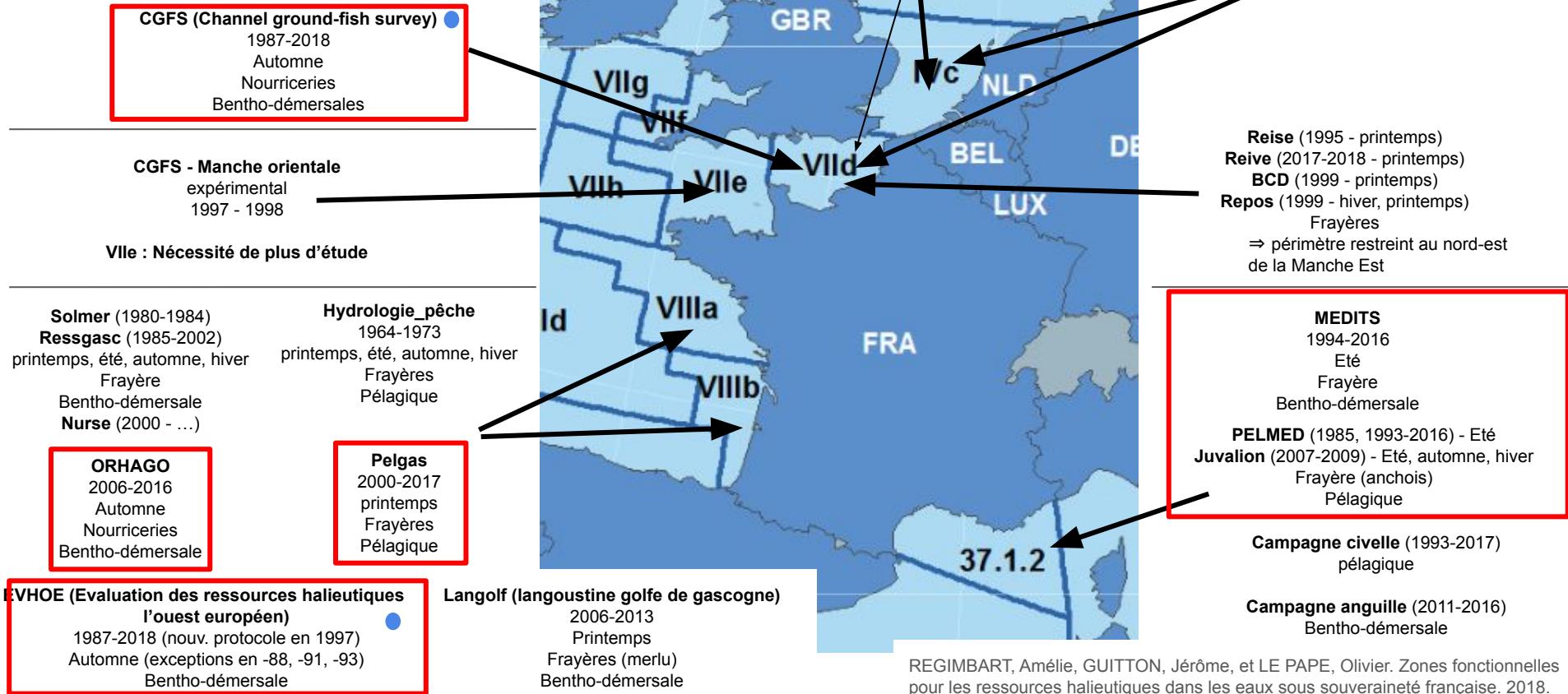


Annexes II

Données (ce qui est dispo et plot)

● Dispo sur DATRAS

Campagnes scientifiques



Stocks d'intérêt

Quelles cas d'étude choisir ?

Des stocks où la donnée est disponible :

- ⇒ stock évaluées,
- ⇒ où les captures sont majoritairement françaises,
- ⇒ où les données VMS sont disponibles (bateaux > 12 m)

Manche-Est/Mer du Nord	Golfe de Gascogne
Sole (<i>Solea solea</i> - Division 27.7.d (Eastern English Channel) Fr (50%) - UK (25%) - Belg. (25%)	Bay of Biscay Sole (<i>Solea solea</i> - sole commune - 8a, 8b) Fr
Cephalopods ???	Nephrops (Divisions 8.a,b, FU 23-24) Fr
Striped red mullet (<i>Mullus surmuletus</i>) - rouget barbet de roche - Subarea 4 (North Sea), divisions 7.d (Eastern English Channel) and 3.a (Skagerrak, Kattegat) Neth (½)- Fr (½) - UK - Belg	European seabass (<i>Dicentrarchus labrax</i> - Bar - Divisions 8.a-b (Bay of Biscay North and Central) Fr
Plaice in Division 7.d Belg. (½) - Fr (½) - UK	Sardine in divisions 8a, b, d Fr (½)- Esp
Whiting (<i>Merlangius merlangus</i> - merlan - Subarea 4 (North Sea), Division 7.d (Eastern English Channel) and 3.a (Skagerrak, Kattegat) Vld : Fr (½) Neth UK Belg.	Hake (<i>Merluccius merluccius</i> - merlu - Division 3.a, Subareas 4, 6 and 7 and Divisions 8.a,b,d (Northern stock) Fr (¼) - Sp - UK - Dan. - Ir.
	Anglerfish (<i>Lophius piscatorius</i> and <i>Lophius budegassa</i> - Baudroie - Sub-area 7 and Divisions and 8.a,b,d) Fr (60%) - UK - Ir - Sp
	Pollack (<i>Pollachius pollachius</i> - lieu jaune - Subarea 8 and Division 9.a) Fr
	Whiting (<i>Merlangius merlangus</i> - merlan - Subarea 8 and Division 9.a) Fr
	Herring - Subarea 8 (Bay of Biscay) - Data poor Fr
	Plaice (<i>Pleuronectes platessa</i> - Plie - Subarea 8 and Division 9.a) Fr (2/3) - Port.

L'utilisation courante de la donnée VMS

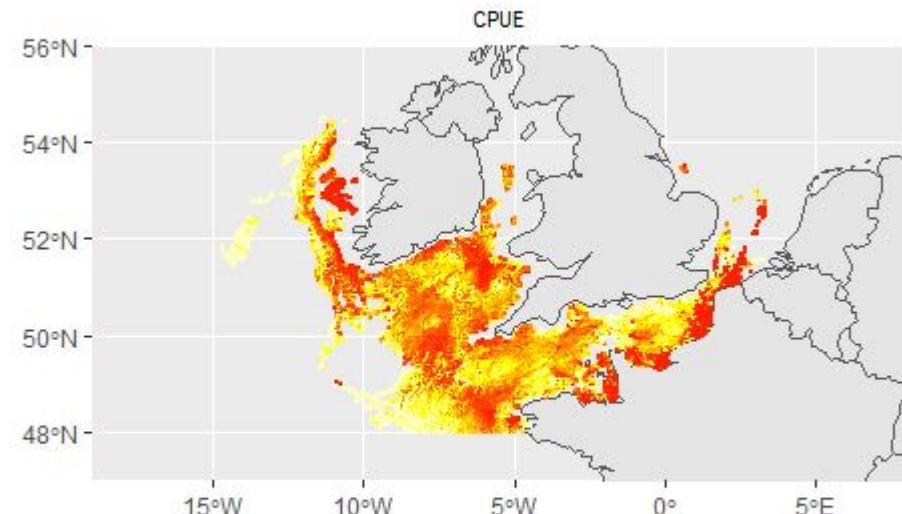
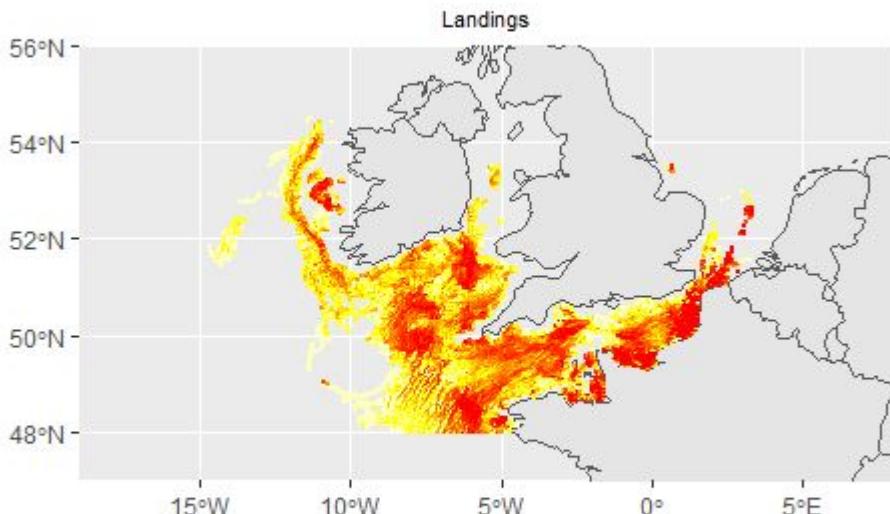
Données VMS

- Distribution de l'effort (Gerritsen et Lordan, 2011)
- Estimer l'impact des engins de pêche sur les habitats (Gerritsen, Minto and Lordan, 2013)
- Analyse du comportement de pêche (Vermard et al., 2010)

Croisement aux données logbooks

- Estimer d'indice d'abondance (Murray et al., 2013 ; Thorson et al., 2016)
- Surveillance de pêche (validation des données des logbooks) (Gerritsen et Lordan, 2011)

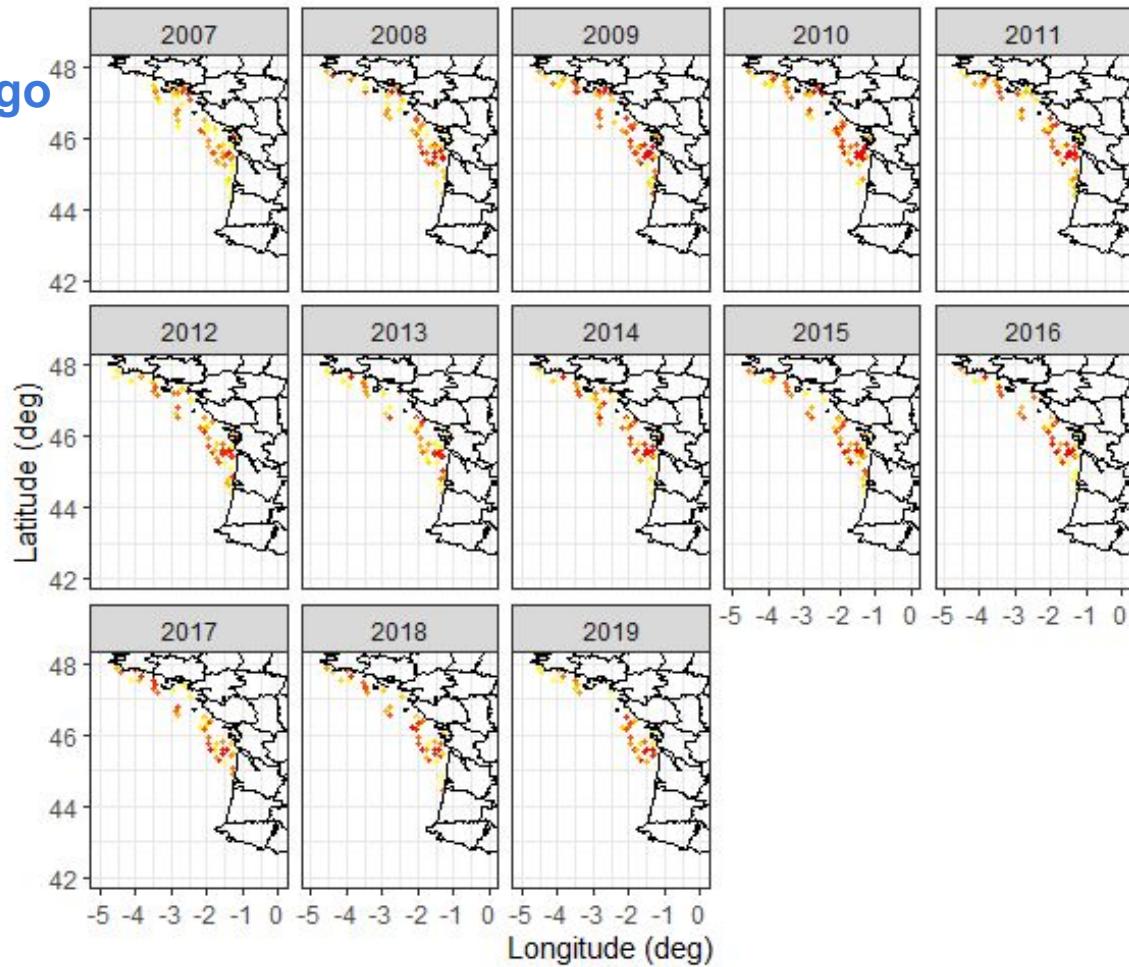
Données VMS (2006) : Solea solea



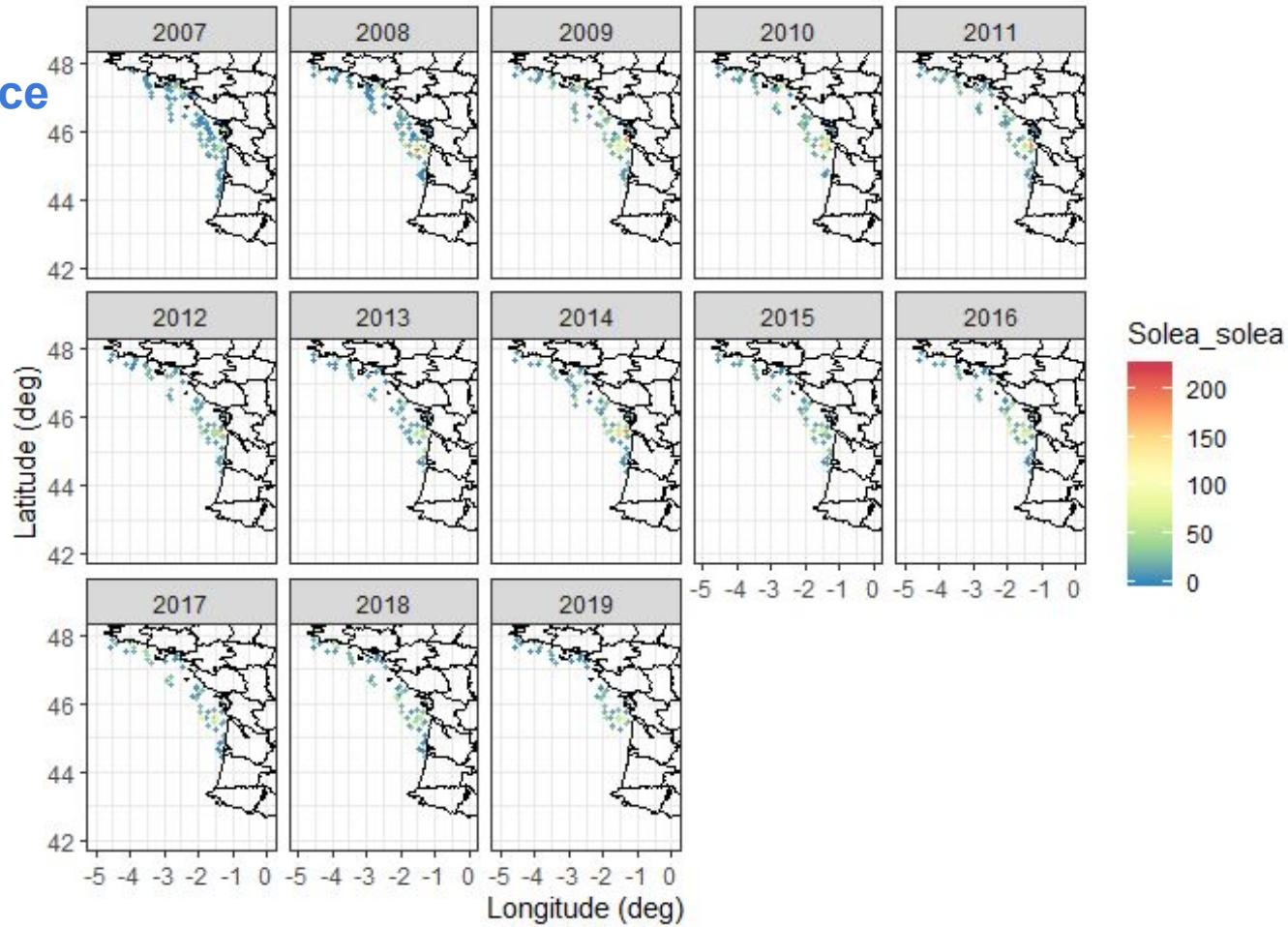
couleur = q10

Données d'abondance Orhago

couleur = q10



Données d'abondance Orhago

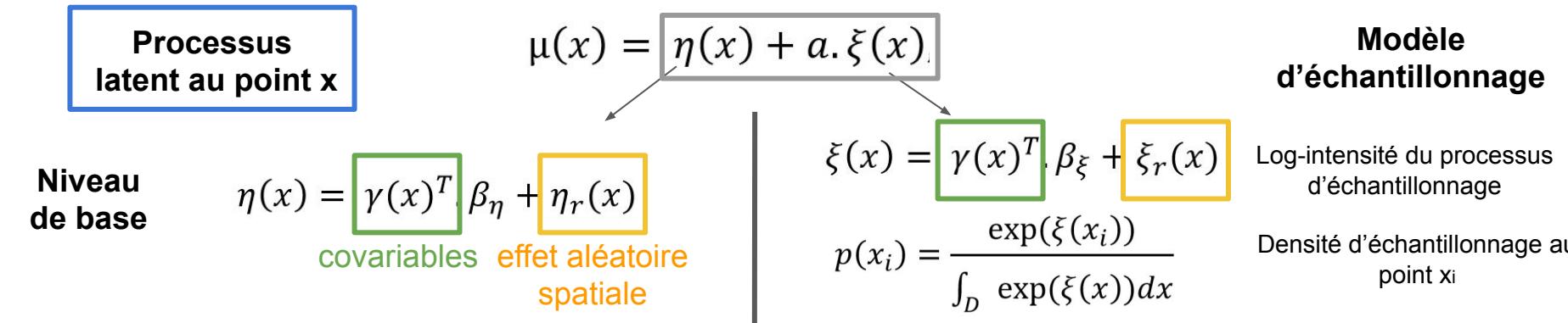


Annexes III

Echantillonnage préférentiel

L'échantillonnage préférentiel : concept de base et modélisation

Modèles de base : Pati et al. (2011)



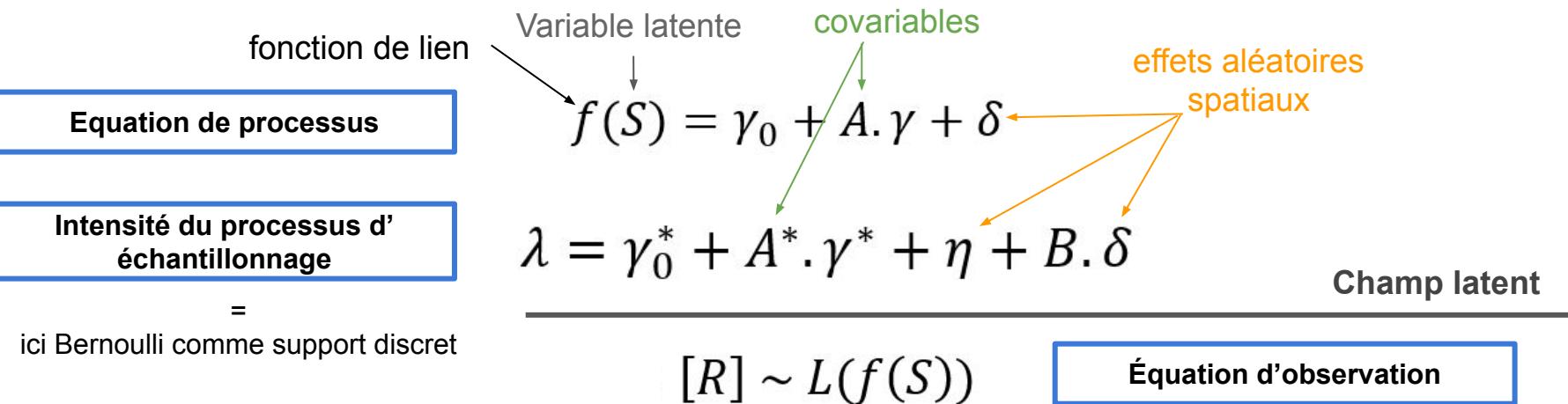
Processus d'observations au point x

$$y_i | x_i \sim N(\eta(x_i) + a \cdot \xi(x_i), \sigma^2)$$

On peut écrire $E(y_i | x_i) = \gamma(x_i)^T \beta^* + a \cdot \xi_r(x_i) + \eta_r(x_i)$ avec $\beta^* = a \cdot \beta_\xi + \beta_\eta$

L'échantillonnage préférentiel : concept de base et modélisation

Modèles de base : Conn et al. (2017)



L'échantillonnage préférentiel : concept de base et modélisation

Approches alternatives :

Méthodes de pondération de la vraisemblance

Irvine et al. (2018), Zidek et al. (2014) ⇒ pondération de la fonction de vraisemblance par l'estimateur de Horvitz-Thompson
(méthode d'échantillonnage à probabilités inégales)

Echantillonnage adaptatif

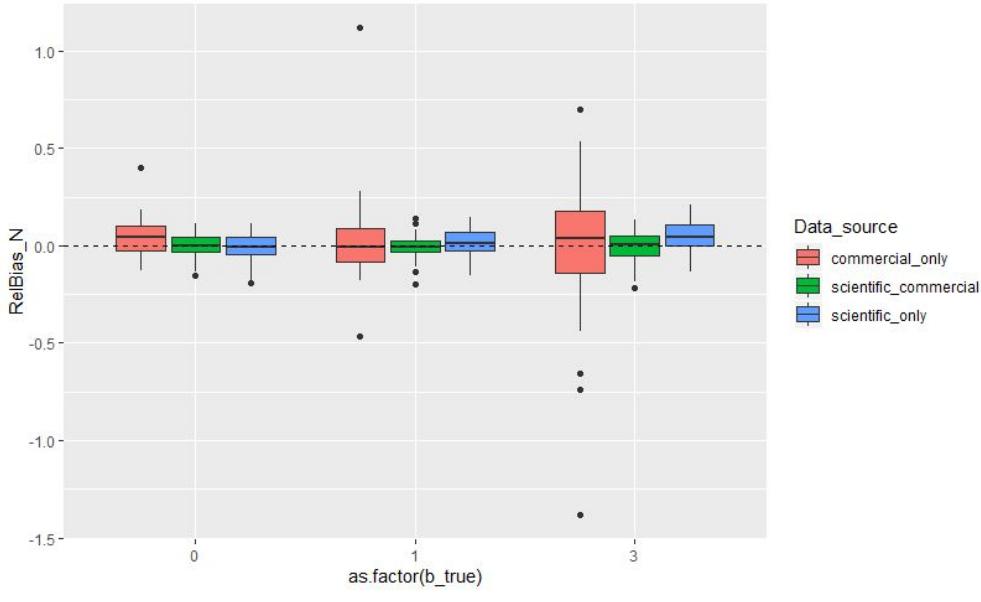
Pacifci et al. (2016)

$$[S_t, X_t, Y_t] \neq [S_t][X_t | S_t][Y_t | S_t, X_t]$$

$$[S_t, X_t, Y_t] = [S_t][X_t | X_{t-1}, Y_{t-1}][Y_t | S_t, X_t]$$

Autres approches

Résultats de simulation



Quand l'échantillonnage est préférentiel

→ estimation de l'abondance n'est pas biaisée mais elle est très imprécise pour la donnée commerciale seule.

→ le fait d'ajouter la donnée scientifique permet de réduire l'incertitude. Mais l'estimation de l'abondance n'est pas plus précise que pour la donnée scientifique seulement.

⇒ **Apport de la donnée commerciale par rapport à la donnée scientifique ?**

Contexte

Sujet de
thèse

Modèle et
hypothèses

Echantillonnage
préférentiel

Résultats de
simulation

Bilan