PLANS D'ÉCHANTILLONNAGE ADAPTÉS AUX DONNÉES SPATIALISÉES D'ACTIVITÉ DE PÊCHE PROFESSIONNELLE ISSUES DE L'OBSERVATOIRE VALPENA[1]

Elodie Plissonneau[1], Lise Bellanger[2], Ion Tillier[3] & Brice Trouillet[4]

[1] *Université de Nantes – Maison des Sciences de l'Homme – USR3491, bureau 505, 5 allée Jacques Berque, BP 12015, 44021 Nantes Cedex 1*

elodie.plissonneau@univ-nantes.fr

[2] *Université de Nantes – Laboratoire de Mathématiques Jean Leray – UMR CNRS 6629, bât.10 bureau 116, UFR Sciences et Techniques, 2 rue de la Houssinière, BP 92208, 44322 Nantes Cedex3*

lise.bellanger@univ-nantes.fr

[3] *COREPEM – Chargé de missions Pêche et Eolien en mer – Comité Régional des pêches maritimes et des élevages marins des Pays de la Loire, 2 rue Colbert, 85100 Les Sables d'Olonne*

ion.tillier@corepem.fr

[4] *Université de Nantes – LETG-Nantes, Géolittomer – UMR CNRS 6554, BP 81227, 44312 Nantes Cedex 3*

brice.trouillet@univ-nantes.fr

Résumé. L'intensification et la diversification des activités en mer impliquent un partage croissant de l'espace maritime. Afin de disposer d'informations sur leurs pratiques, les marins pêcheurs professionnels avec des universitaires ont développé une démarche originale basée sur une cartographie fine des activités de pêche. Reposant sur des enquêtes déployées chaque année auprès de la totalité des marins pêcheurs, diverses contraintes jouent sur la pérennité du dispositif de recueil des données. Afin de conserver un niveau de précision élevé tout en limitant le nombre d'enquêtes, différentes stratégies d'échantillonnage ont donc été créées, en se basant sur la théorie des sondages (calcul du nombre optimal de navires à enquêter, allocation optimale/proportionnelle, etc) et comparées entre elles. Les simulations effectuées mettront en évidence qu'un plan d'échantillonnage stratifié aléatoire simple appliqué aux bases Navires s'avère adapté à l'estimation spatialisée d'indicateurs.

Mots-clés. Théorie des sondages, plan d'échantillonnage, stratification, allocation, estimation spatialisée, pêche maritime, dynamiques spatio-temporelles, aménagement de l'espace maritime.

[1] éVALuation des activités de Pêche en regard des Nouvelles Activités (VALPENA), détails sur valpena.fr

1 Introduction

L'intensification des activités en mer, mais aussi leur diversification (émergence des énergies marines renouvelables, des aires marines protégées, de l'aquaculture offshore, etc.), contraignent les marins pêcheurs professionnels à modifier leurs pratiques et/ou leurs espaces de travail. Ces activités créent une pression de plus en plus importante sur la pêche maritime professionnelle et peuvent être à l'origine de conflits de plus en plus nombreux.

Dans ce contexte, la spatialisation des activités de pêche est devenue un enjeu de connaissance capital. La création d'un système d'observation spatio-temporelle normalisée des activités de pêche était indispensable. À cet effet, le projet VALPENA (éVALuation des activités de PÊche au regard des Nouvelles Activités) a vu le jour en 2010 et est devenu un Groupement d'Intérêt Scientifique (GIS) en 2014. Il est issu d'une co-construction entre scientifiques et structures professionnelles du domaine des pêches et est conçu comme une plateforme collaborative structurant des observatoires régionaux des pratiques de pêche. Aujourd'hui, il réunit 7[2] des 9 CRPMEM[3] à l'échelle nationale métropolitaine (soit environ 3000 navires sur les 4500 de la flotte métropolitaine). Les données issues des observatoires VALPENA sont exploitées par les structures professionnelles des pêches et servent de socle opérationnel à la défense de l'intérêt des marins pêcheurs.

Les observatoires VALPENA rencontrent cependant deux problèmes majeurs dans la mise en œuvre des enquêtes auprès des pêcheurs :

- le manque de moyens des comités pour enquêter la totalité des navires chaque année ;
- le refus des pêcheurs d'être enquêtés, lié à la sur-enquête dont ils font l'objet.

Pour pallier ces deux problèmes, il a été décidé de mettre en place un plan d'échantillonnage (PE). Dans un premier temps, un premier PE (*PEactuel*) a été construit. Il conduit à n'enquêter qu'un tiers de la flotte chaque année, de manière à couvrir l'intégralité de la flotte au bout de trois années. Il a été mis en place au sein du COREPEM[4] pour les enquêtes réalisées de 2013 à 2015, sur l'activité de l'année précédente. Le but de cette présentation est de comparer le *PEactuel* à d'autres utilisant la théorie des sondages, pour aboutir à des recommandations de mise en œuvre sur le terrain. La comparaison est effectuée à l'aide de l'Indicateur d'Intensité (*II*) sur des plans d'échantillonnage avec des allocations proportionnelles ou optimales.

2 Matériel et Méthodes

2.1 Matériel

Les données à traiter concernent aujourd'hui les deux-tiers de la flotte française métropolitaine. Avec une clé qui est le navire, les bases de données VALPENA combinent :

- des données d'enquête localisant, à l'échelle mensuelle, les activités de pêche en fonction d'un référentiel spatial constitué d'un maillage géométrique de 3 milles nautiques de côté, en distinguant engins de pêche mis en œuvre et espèces ciblées ;
- des données techniques et administratives sur les navires de pêche : port d'attache et d'exploitation, caractéristiques techniques (longueur, puissance, etc.), armateurs, etc.

Ce système permet aujourd'hui de fournir des données capitales et inédites, tant pour que les pêcheurs puissent exprimer leurs intérêts qu'à des fins de recherche. Elles complètent d'autres données détenues par le Ministère en charge de la pêche. Différents travaux permettent de

[2] Nord-Pas-de-Calais/Picardie, Haute et Basse Normandie, Bretagne, Pays de la Loire, PACA et Poitou-Charentes.
[3] Comités Régionaux des Pêches Maritimes et des Élevages Marins.
[4] Comité Régional des Pêches maritimes et des Élevages Marins des Pays de Loire.

comprendre le format des données ainsi que la méthode déployée pour les collecter et les traiter (Plissonneau (2014), Tillier et al. (2014) ou Trouillet (2013)).

Afin d'éviter d'enquêter tous les navires tous les ans, tout en maintenant un bon niveau de représentativité, un travail sur le recueil des données utilisant un PE a été initié. Différents PE sont comparés à l'aide de simulations ; afin d'améliorer la représentativité grâce à l'utilisation de la théorie des sondages. Les résultats sont présentés ici pour les données de l'activité de l'année 2010 de la flotte de Pays de la Loire.

2.2 Méthodes

2.2.1 Indicateur

La population étudiée est la population de navires (i.e. la flottille étudiée) et sera noté dans la suite U. Les navires sont les individus de cette population et seront notés nav_i, pour $i \in \{1, \dots, N\}$. De plus, on a : MO le nombre de mois de la période étudiée et S le nombre de mailles totales fréquentée par toute la population.

Afin de construire et comparer les différents PE, en concertation avec les différents comités membres du GIS, le paramètre d'intérêt sera l'Indicateur d'Intensité (II), un des 7 indicateurs utilisés dans la méthodologie du GIS.

On définit l'indicateur d'intensité pour le navire nav_i et la maille s comme le nombre total de mois travaillés sur cette maille pour ce navire i.e. :

$$II_{nav_i}^{s} = \sum_{m=1}^{MO} 1_{mois_m \times maille_s \times nav_i}.$$

Dans la suite, notre variable d'intérêt Y_i correspond à cet indicateur moyenné sur l'ensemble des mailles, i.e. :

$$Y_i = \frac{\sum_{s=1}^{S} II_{nav_i}^{s}}{S}.$$

Le paramètre que l'on cherche à optimiser sera la moyenne de cette variable d'intérêt sur tous les navires

$$\bar{Y}_{pop} = \frac{\sum_{i=1}^{N} Y_i}{N}.$$

2.2.2 Plans d'Échantillonnage

Afin de tester l'efficacité du *PEactuel*, PE mis en place au sein des Pays de la Loire de 2013 à 2015, et de l'améliorer, 4 autres PE stratifiés aléatoires simples avec allocation proportionnelle ou optimale (notés par la suite *PEstrat*, *PEstratn, PEstratopt et PEstratnopt*) ont été créés grâce aux techniques de la théorie des sondages (voir par exemple les ouvrages de référence de Ardilly (2006), Lumley (2010) ou Tillé (2001)) et comparés par simulation. Tous ces PE sont des PE aléatoires simples stratifiés.

Le but de ces simulations est d'étudier les améliorations possibles à apporter au PE actuel pour conserver une bonne précision en limitant au maximum la complexité de traitement des données, effectué au sein des comités. En effet, plus la sélection des navires sera compliquée, plus les calculs d'indicateurs seront complexes, notamment par l'ajout d'un poids sur chacun des navires, accentué avec l'allocation optimale. Différents taux d'enquête, différentes stratifications mais aussi

différentes allocations (allocation optimale et allocation optimale de Neyman) ont donc été testés.

i. Types d'allocations

Pour calculer le nombre d'unités à enquêter dans chaque strate, il existe deux principaux types d'allocation : l'allocation proportionnelle et l'allocation optimale.

L'allocation proportionnelle, méthode simple à mettre en œuvre mais aussi à expliquer a d'abord été envisagée. Un plan de sondage stratifié est dit à allocation proportionnelle si

$$\frac{n_h}{N_h} = \frac{n}{N}$$

où :

- n_h est la taille de l'échantillon de la strate h ;
- N_h est la taille de la strate h ;
- n est la taille de l'échantillon à prélever sur la population totale ; dans nos simulations, il correspond soit au tiers de la population étudiée, soit au nombre optimal de navires à enquêter pour une précision de 10% calculé à partir de la théorie des sondages ;
- N est la taille de la population totale.

L'utilisation de cette méthode est cependant contestée, notamment parce qu'elle ne permet pas l'exploitation optimale de la stratification proposée. Afin de pallier ce problème, et ainsi améliorer la précision de l'estimateur, nous avons décidé de tester des stratégies d'échantillonnage basées sur l'allocation optimale de Neyman. Celle-ci tient compte de la dispersion S_h de la variable d'intérêt y :

$$n_h = n\frac{N_h S_h}{\sum_{l=1}^{H} N_l S_l}, h = 1, \dots, H$$

où

- $S_h = \sqrt{S_h^2}$
- $S_h^2 = \frac{1}{N_h - 1}\sum_{k \in U_h}(Y_k - \bar{Y}_h)^2$ est la variance corrigée de la strate h ;
- U_h est la population de la strate h ;
- $\bar{Y}_h = \frac{\sum_{k \in U_h} Y_k}{N_h}$ est la moyenne des Y_k pour $k \in U_h$.

Le principe est de tirer une proportion d'individus (ie les navires) à enquêter d'autant plus importante que la variance au sein de la state est grande. Cette méthode a cependant l'inconvénient de complexifier les calculs, en ajoutant des poids aux navires en fonction de leur probabilité d'inclusion, ce qui n'est pas vraiment envisageable dans le cadre de la méthodologie de traitements des données VALPENA.
On comparera donc à l'aide de simulations ces deux stratégies d'allocation afin de tester leur efficacité.

ii. Les différents PE étudiés

Cinq plans d'échantillonnage ont été comparés pour les données d'activité de 2010 pour la région Pays de la Loire. Leurs descriptifs sont les suivants :

PEactuel
Taille de l'échantillon : 1/3 de la flotte enquêtée ;
Stratification : basée sur le type d'engin utilisé et le quartier maritime d'immatriculation ;
Allocation proportionnelle ;
Objectif : avoir parcouru l'ensemble de l'échantillon en 3 ans.

PEstrat
Taille de l'échantillon : 1/3 de la flotte enquêtée ;
Stratification : affinée grâce au savoir des comités ;
Allocation proportionnelle ;
Objectif : représentativité excellente + tirage aléatoire chaque année.

PEstratopt
Taille de l'échantillon : 1/3 de la flotte enquêtée ;
Stratification : affinée grâce au savoir des comités ;
Allocation optimale ;
Objectif : représentativité excellente + tirage aléatoire chaque année.

PEstratn
Taille de l'échantillon : optimisée en utilisant la théorie des sondages (Tillé, 2001, chapitres 4 et 7) ;
Stratification : identique à celle du *PEstrat* ;
Allocation proportionnelle ;
Objectif : représentativité bonne (précision 5 à 10%) avec un minimum de navires (calculs plus simples que PEstratnopt) + tirage aléatoire chaque année.

PEstratnopt
Taille de l'échantillon : optimisée en utilisant la théorie des sondages (Tillé, 2001, chapitres 4 et 7) ;
Stratification : identique à celle du *PEstrat* ;
Allocation optimale ;
Objectif : représentativité excellente (meilleure que PEstratn) avec un minimum de navires + tirage aléatoire chaque année.

La construction des *PEactuel*, *PEstrat* et *PEstratn* est détaillée dans Plissonneau (2014). Celle des *PEstratopt* et *PEstratnopt* sont détaillées dans Plissonneau (2016). La comparaison par simulation de ces différents types d'allocation va permettre d'étudier si l'amélioration de la précision de l'estimateur apportée par l'allocation optimale semble suffisante par rapport à la complexification des calculs qu'elle implique.

2.2.3 Données et Méthode de comparaison des PE

Les données utilisées seront celles de l'activité des navires de Pays de la Loire en 2010.
Afin de déterminer le PE qui représente le plus précisément l'activité de la flotte totale, l'II_p associé à chaque maille h a été calculé pour chacun des 5 PE et comparé à la valeur obtenue pour la flotte totale.

La méthode de simulation est la suivante :

- Calcul de $\bar{Y}_U$ pour chacune des mailles pour l'ensemble de la flotte ;
- Pour chacun des 5 PE, pour $b = 1, \dots, 500$:
    - tirage aléatoire d'un échantillon b;
    - calcul de l'$\bar{Y}_b$;
- Comparaison des résultats entre $\bar{Y}_U$ et $\bar{Y}_b$: calcul des biais, variances et Erreur Quadratique Moyenne (EQM) absolus et relatifs.

3 Résultats et perspectives

Nous montrerons que le nouveau plan d'échantillonnage stratifié proposé permet d'enquêter un nombre de navires plus faible que *PEactuel* mis en place en 2013, 2014, 2015 en Pays de la Loire, avec une précision meilleure, tout en respectant un tirage totalement aléatoire.

Des simulations ont aussi été réalisées pour les régions Bretagne, Nord-Pas-de-Calais/Picardie et Basse Normandie aboutissant aux mêmes conclusions que pour les Pays de la Loire. Grâce à la mise en place d'une méthode de tirage aléatoire suivant la technique des sondages aléatoires simples stratifiés, une liste de navires à enquêter en 2015 et 2016 a pu être produite. Cette approche sera généralisée à l'ensemble des régions ayant adopté le dispositif VALPENA, de manière à proposer à l'ensemble des comités enquêteurs une liste spécifique de navires à enquêter.

4 Avertissement

L'ensemble de ces travaux est réalisé dans le cadre des travaux du GIS VALPENA, issu d'un projet financé par le SMIDAP (Conseil Régional des Pays de la Loire) de 2010 à 2012, ainsi qu'avec le soutien du projet interdisciplinaire du CNRS - Défi Littoral 2016 ESPADON "Echantillonnage SPAtialisé pour la collecte de DONnées : application à la pêche maritime professionnelle". Nous remercions le CRPMEM Pays de la Loire d'avoir permis l'utilisation de leurs données VALPENA issues de leur campagne d'enquêtes 2011 pour illustrer notre présentation. Nous remercions également Guillaume Chauvet (enseignant chercheur à l'ENSAI) pour ses précieux conseils.

Bibliographie

[1] Ardilly, P. (2006), *Les techniques de sondage (nouvelle édition)*, Editions Technip, Paris.

[2] Lumley, T. (2010), *Complex Surveys: A Guide to Analysis Using R*, Wiley, Hoboken, NJ, USA.

[3] Plissonneau, E. (2014), Normalisation de l'exploitation des données VALPENA par la mise en place d'indicateurs géostatistiques et validation de plans d'échantillonnage avec la théorie des sondages, Rapport de stage de Master 2, 66p.

[4] Plissonneau, E. (2016), Plans d'Échantillonnage, document interne du GIS VALPENA

[5] Tillé, Y. (2001), *Théorie des sondages – Échantillonnage et estimation en populations finies,* Dunod, Paris.

[6] Tillier, I., Cautain, B., Brivoal, F. and N. Gace-Rimaud (2014), Etude d'impact du raccordement électrique du projet éolien du banc de Guérande : Analyse de l'activité de pêche professionnelle, Rapport produit dans le cadre de l'étude d'impact du projet, Réseau de Transport d'Electricité, 99p.

[7] Trouillet, B. (2013), Connaissance des pratiques spatiales dans le domaine des pêches maritimes : l'expérience VALPENA. Colloque « Exploitation et conservation des écosystèmes aquatiques : une question d'échelle ? », AFH, 19-21 juin 2013, Bordeaux.