

Construction des indicateurs socio-économiques sur la pêche professionnelle en France métropole : partie 1 - plan d'échantillonnage



Fiche documentaire

Titre du rapport : Construction des indicateurs socio-économiques sur la pêche professionnelle en France métropole : partie 1 - plan d'échantillonnage	
Référence interne : R.DEP/UNIT/LABO AN- NUM	Date de publication : AAA/MM/JJ Version : 1.0.0
Diffusion : <input checked="" type="checkbox"/> libre (internet) <input type="checkbox"/> restreinte (intranet) – date de levée d'embargo : AAA/MM/JJ <input type="checkbox"/> interdite (confidentielle) – date de levée de confidentialité : AAA/MM/JJ	Référence de l'illustration de couverture Crédit photo/titre/date Langue(s) :
Résumé/ Abstract : Dans le cadre du règlement européen DCF (Data Collection Framework), les états membres ont pour obligation de collecter et fournir des données socio-économiques concernant la pêche, pour servir de base de conseils sur la gestion des pêches ou pour d'autres usages. En France métropole, le producteur de données est le Service de la Statistique et de la Prospective (SSP) du ministère de l'Agriculture et de l'Alimentation. Le SSP s'appuie sur ses deux partenaires pour la collecte des données : l'Ifremer qui met en œuvre des enquêtes sur le terrain et le Laboratoire d'Économie et de Management de Nantes-Atlantique (Lemna) qui collecte des données comptables auprès des centres de gestion. Ce document présente les étapes de construction du plan d'échantillonnage annuel qui sera utilisé ensuite pour la collecte des données économiques. Il s'agit de la première étape de la construction des indicateurs socio-économiques, précédant la collecte de données et la validation des données.	
Mots-clés/ Key words : Pêche professionnelle, enquêtes économiques, socio-économie, indicateurs économiques, plan d'échantillonnage, collecte, validation	
Comment citer ce document : Merzéréaud M., Daurès F., Guyader O., Le Grand C., Leonardi S., Macher C., Spagnol C. 2021. Construction des indicateurs socio-économiques sur la pêche professionnelle en France métropole : partie 1 - plan d'échantillonnage, Rapport Ifremer-RBE-EM-SIH	
Disponibilité des données de la recherche :	
DOI :	

Commanditaire du rapport :	
Nom / référence du contrat :	
<input type="checkbox"/> Rapport intermédiaire (réf. bibliographique : XXX) <input type="checkbox"/> Rapport définitif (réf. interne du rapport intermédiaire : R.DEP/UNIT/LABO AN- NUM/ID ARCHIMER)	
Projets dans lesquels ce rapport s'inscrit (programme européen, campagne, etc.) :	
Auteur(s) / adresse mail	Affiliation / Direction / Service, laboratoire
Merzéréaud Mathieu	RBE/EM
Daurès Fabienne	RBE/EM
Guyader Olivier	RBE/EM
Le Grand Christelle	RBE/EM
Leonardi Sophie	RBE/EM
Macher Claire	RBE/EM
Spagnol Charlène	RBE/STH/LBH
Contributeur(s) / adresse mail	Affiliation / Direction / Service, laboratoire
Minne Marie-Dominique	SSP, Ministère de l'Agriculture et de l'Alimentation
Encadrement(s) :	
Destinataire :	
Validé par :	

Sommaire

Table des matières

1	Introduction	6
2	Construction du plan d'échantillonnage	7
2.1	Constitution de la base de sondage	7
2.1.1	Etape n°1 : Mobilisation des deux sources de données utilisées : le fichier flotte FPC et les données d'activité des navires de pêche	7
2.1.2	Etape n°2 : Construction de segmentations adaptées à différents besoins de restitution d'indicateurs	8
2.1.3	Etape n°3 : Production d'indicateurs individuels servant de critères d'inclusion dans la base de sondage	8
2.1.4	Composition de la base de sondage résultante	9
2.1.5	En résumé	11
2.2	Calculs des paramètres d'ajustement du chiffre d'affaires annuel individuel par la variable composite FILEMO	11
2.2.1	Présentation de la méthode	11
2.2.2	Synthèse de l'analyse des résultats obtenus : application AT	12
2.3	Méthodologie d'allocation appliquée au plan d'échantillonnage : problème d'optimisation de Neyman sous contraintes	13
2.4	Tirage de l'échantillon et répartition entre partenaires de collecte	14
3	Références bibliographiques	17

1 Introduction

Dans le cadre du règlement européen (UE) 2017/1004 dit règlement DCF pour « Data Collection Framework »¹, les états membres ont pour obligation de collecter et fournir des données socio-économiques concernant la pêche, définies dans la décision déléguée (UE) 2021/1167 de la Commission du 27 avril 2021², pour servir de base de conseils sur la gestion des pêches ou pour d'autres usages.

En France métropole, le producteur de données est le Service de la Statistiques et de la Prospective (SSP) du ministère de l'Agriculture et de l'Alimentation. Le SSP s'appuie sur ses deux partenaires pour la collecte des données : l'Ifremer qui met en œuvre des enquêtes sur le terrain et le Laboratoire d'Économie et de Management de Nantes-Atlantique (Lemna) qui collecte des données comptables auprès des centres de gestion.

Pour produire ces données, des enquêtes sont menées chaque année auprès des armements, dans le but de calculer des indicateurs économiques par segment de flotte et plus généralement, d'évaluer les performances économiques des différentes flottilles et leur évolution. Les principaux thèmes abordés lors des enquêtes sont : les revenus, les coûts, la valeur du capital physique, les investissements matériels et la situation financière, et les emplois. Cette enquête est labellisée d'intérêt général et de qualité statistique par le Conseil National de l'Information Statistique (CNIS) depuis 2012³. Ce label permet notamment d'assurer une méthodologie rigoureuse et un respect strict de la confidentialité des données, protégées par le secret statistique.

Ce document présente les étapes de construction du plan d'échantillonnage annuel qui sera utilisé ensuite pour la collecte des données économiques, par les deux partenaires de collecte du SSP : l'Ifremer et le Lemna. Il présente donc la première étape de la construction des indicateurs socio-économiques et il est à mettre en relation avec les documents suivants :

Spagnol C., Le Grand C., Guyader O., 2021. Construction des indicateurs socio-économiques sur la pêche professionnelle en France métropole : partie 2 - collecte de données

Le Grand C., Daurès F., Guyader O., Macher C., Leonardi S., Merzereaud M. 2021. Construction des indicateurs socio-économiques sur la pêche professionnelle en France métropole : partie 3 - méthode de validation

¹ Règlement (UE) 2017/1004 du Parlement Européen et du conseil du 17 mai 2017 : <https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32017R1004&from=EN>

² Voir tableau 7 de la décision déléguée (UE) 2021/1167 de la Commission du 27 avril 2021 : <https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32021D1167&rid=2>

³ Production de données économiques dans le secteur des pêches maritimes (Enquête pour la (2021A704AG) - <https://www.cnis.fr/enquetes/enquete-pour-la-production-de-donnees-economiques-dans-le-secteur-des-peches-maritimes-2021a704ag/>

2 Construction du plan d'échantillonnage

La construction du plan d'échantillonnage annuel pour la collecte de données économiques sur le secteur de la pêche dans le cadre du programme national en métropole est sous la responsabilité du Service de la Statistique et de la Prospective (SSP) et implique les partenaires de collecte : l'Ifremer et le Lemna. L'Ifremer est en charge des deux premières phases de la construction du plan d'échantillonnage, à savoir la constitution de la base de sondage ainsi que le calcul des paramètres d'ajustement du chiffre d'affaires annuel individuel par la variable composite FILEMO. Il est également proposé dans cette partie une description simplifiée des phases consécutives mises en œuvre par le SSP (calcul d'allocation d'enquêtes, tirage aléatoire de l'échantillon et répartition de la collecte entre les deux partenaires), permettant ainsi au lecteur de disposer d'une vue d'ensemble du processus.

2.1 Constitution de la base de sondage

Tout plan d'échantillonnage nécessite une **base de sondage** décrivant la population étudiée et au sein de laquelle seront tirées les unités à enquêter. Cette base a également vocation à intégrer un ensemble d'informations individuelles, qui permettront de répondre aux besoins inhérents aux différentes phases opérationnelles planifiées, allant du tirage aléatoire jusqu'à l'estimation des indicateurs finaux (en passant par la constitution de la stratification, les calculs d'allocation basés sur des objectifs de précision prédéterminés, des procédures d'inférence statistique telles que le calage sur marges, ...).

Dans cette partie seront décrites les étapes de construction de la base de sondage utilisée dans le cadre du plan d'échantillonnage pour la collecte des données économiques en France métropolitaine. **Deux versions distinctes** de cette base seront créées, une pour la **façade Manche-Mer du Nord-Atlantique** (qu'on distinguera par l'abréviation « AT » dans la suite du document, et couvrant les navires opérant en zone FAO 27), une autre décrivant les navires basés en **Méditerranée** (qu'on distinguera par l'abréviation « ME » dans la suite du document, affectés à la zone FAO 37). La construction de ces deux bases se fonde sur une **méthodologie commune**, leur différenciation reposant simplement sur la population considérée ainsi que sur une caractérisation spécifique de la typologie employée. Dans cette description, on prendra pour **référence temporelle l'année N de mise en œuvre du processus de collecte**, sachant que l'enquête s'attachera à collecter les données portant sur l'année N-1.

2.1.1 Etape n°1 : Mobilisation des deux sources de données utilisées : le fichier flotte FPC et les données d'activité des navires de pêche

Les deux bases de sondage sont issues des mêmes sources de données individuelles exhaustives. Elles prennent en effet majoritairement appui sur le **fichier national de la Flotte de Pêche Communautaire (FPC)**, fichier compilant des informations administratives et des caractéristiques techniques relatives à chaque navire de la flotte de pêche française. Ce fichier sera enrichi d'une typologie avancée mise en place grâce au **recensement exhaustif de l'activité des navires** par le biais d'un travail de reconstitution des calendriers d'activité⁴. Afin de s'assurer de la cohérence des informations contenues dans la base de sondage (constituée en fin d'année N-1) avec la période ciblée par l'enquête (pour rappel, elle aussi l'année N-1), une version actualisée de ce fichier est utilisée. Toutefois, une seconde version de ce fichier, datée au 31 décembre de l'année

⁴ <https://sih.ifremer.fr/Activite-socio-economie/Activite-des-navires>

N-2, est mise à contribution de façon complémentaire sur deux aspects spécifiques. D'une part, elle intègre **une typologie des flottilles validée et exhaustive, la plus actualisée** qui puisse être mise à disposition au moment de la construction de cette base. D'autre part, elle propose une photographie plus ancienne qui, mise en parallèle avec l'image procurée par le fichier actualisé, permet **d'isoler les navires présents tout au long de l'année N-1**, tout comme ceux qui auront connu durant cette année des changements de situation notables (on verra plus loin que cela constituera des critères d'exclusion de la base de sondage).

2.1.2 Etape n°2 : Construction de segmentations adaptées à différents besoins de restitution d'indicateurs

Une segmentation est une manière de compartimenter une population de navires en « segments », **un segment étant la combinaison d'une flottille d'appartenance⁵ et d'une classe de taille** de navire. On construit, à partir des typologies issues des calendriers d'activité au 31/12 de l'année N-2, des segmentations sur lesquelles reposeront les stratifications structurant le futur plan d'échantillonnage. La finalité de ce dernier étant de construire des indicateurs socio-économiques en réponse aux besoins de la DCF, mais également aux besoins des travaux de recherche et d'expertise de l'Ifremer menés à différentes échelles, il est nécessaire de prendre en compte dans le processus de construction, des **stratifications** répondant à ces **différents niveaux de restitution attendus**. Ainsi, à chacune des deux façades correspondra un découpage spécifique des classes de taille des navires, auquel on associera deux typologies de flottilles distinctes (une **typologie « DCF »** conforme aux nomenclatures requises par les instances européennes, et une **typologie « SIH »** correspondant davantage au besoin scientifique) pour aboutir finalement à deux segmentations par façade⁶.

2.1.3 Etape n°3 : Production d'indicateurs individuels servant de critères d'inclusion dans la base de sondage

Comme énoncé précédemment, la mise en parallèle de deux photographies d'une même base d'information prises aux instants N-2 et N-1 va permettre la construction de multiples **indicateurs individuels de changements d'état**, témoignant de la **stabilité** de l'activité de pêche et du **mode de gestion** des navires considérés durant la période ciblée par l'enquête économique. Cette enquête portant sur l'activité opérée au cours d'une année entière, cette notion de stabilité appliquée à différents facteurs significatifs va naturellement discriminer les navires que l'on considèrera comme pertinents à enquêter, et qui composeront ainsi notre base de sondage. Ainsi, l'intégration dans la base sera conditionnée par la combinaison de 5 critères binaires distincts :

5 Une typologie de flottille a pour but de regrouper des navires ayant des comportements ou des stratégies de pêche similaires (mêmes métiers ou combinaisons de métiers), ou selon des critères tels que les caractéristiques techniques des navires ou les secteurs de provenance ou d'activité.

6 La typologie des flottilles DCF est basée sur l'utilisation d'un engin dominant à l'année tandis que la typologie Ifremer SIH est basée sur la polyvalence des navires du point de vue des engins utilisés (IFREMER SIH (2022). Activité des navires . IFREMER <https://doi.org/10.12770/d31aac89-5d6b-4bee-9e92-611579a846d2>)(Berthou Patrick, Daures Fabienne, Guyader Olivier, Leblond Emilie, Merrien Claude, Demaneche Sebastien, Jezequel Michele (2003). Typologies des flottes de pêche : Méthodes Ifremer-SIH. DRV/SIH/N°4/082003. <https://archimer.ifremer.fr/doc/00705/81686/>). À titre d'exemple, un navire pratiquant 5 mois la drague à coquille-St-Jacques et 7 mois le chalut de fond à poissons sera affecté à la flottille DCF des senneurs et chalutiers de fond et aux flottille / sous-flottille Ifremer des chalutiers non exclusifs / chalutiers dragueurs.

- indicateur de **présence en flotte au 31/12 de l'année N-2** (équivalent à la présence d'une valeur de typologie de flottilles renseignée) : si 0, le navire n'aura pas couvert entièrement la « période » N-1 (exclusion) ;
 - indicateur de **sortie de flotte** en cours d'année N-1 (équivalent à la présence dans le fichier FPC d'une date de référence antérieure à la date de référence majoritaire correspondant à la dernière actualisation de la base) : si 1, le navire est sorti de flotte avant l'actualisation (exclusion) ;
 - indicateur de **changement d'armateur** au cours de l'année N-1 (équivalent à un changement de valeur du champ Armateur entre les versions N-2 et N-1 des sources de données) : si 1, exclusion de la base de sondage ;
 - indicateur de **changement de façade** au cours de l'année N-1 (équivalent à un changement de valeur du champ Façade entre les versions N-2 et N-1 des sources de données) : si 1, exclusion de la base de sondage ;
 - indicateur de navires **inactifs** au 31/12 de l'année N-2 (équivalent à la présence de la modalité « Inactifs à la pêche » dans les champs relevant des typologies de flottille⁷) : si 1, exclusion de la base de sondage.

Il faut noter, concernant ce dernier indicateur d'inactivité, qu'un état des lieux plus actualisé serait bien entendu souhaitable afin de préserver au maximum la base de sondage de navires ayant été éventuellement inactifs durant l'année d'intérêt N-1. Mais comme dit précédemment, cette information actualisée ne peut être obtenue au moment de la constitution du plan d'échantillonnage, ni même au moment de la campagne de collecte. Ce décalage peut représenter un léger biais dans le fait de potentiellement exclure de la base de sondage des navires ayant pourtant été bel et bien actifs durant l'année N-1, tout comme dans le fait d'y avoir peut-être inscrit des navires étant devenus inactifs cette année-là. Une analyse a toutefois été menée à propos de la probabilité d'un navire de redevenir actif après avoir été inactif l'année précédente. Celle-ci montre une **forte tendance du statut d'inactif à perdurer dans le temps**, ce qui peut laisser supposer **un biais minimisé**.

De ces 5 indicateurs, on déduit finalement l'appartenance ou non d'un navire à la base de sondage : on considèrera ainsi qu'un navire peut y être intégré s'il était présent en flotte au 31/12 de l'année N-2 et qu'il l'était toujours selon les dernières informations disponibles, dans la même façade et avec le même armateur, et qu'il était en activité selon les dernières sources d'information disponibles.

2.1.4 Composition de la base de sondage résultante

Nous présentons ici le Tableau 1 **synthétique** décrivant les **champs** constituant la **base de sondage** telle que fournie au SSP. Cette base de sondage servira de support de référence aux différentes étapes de mise au point de la campagne de collecte.

⁷ Sont également compris dans les « inactifs à la pêche » les navires pratiquant la pêche des vers (dragage ou plongée).

Tableau 1 : Champs constitutifs de la base de sondage

CHAMPS	DESCRIPTION FR
NAVS_COD	Code d'immatriculation du navire
AN_REF	Année de référence
NAVP_LONGUEUR_HT	Longueur Hors Tout (en cm)
LONGUEUR_CLASS_LIB	Classe de longueur
QAM_LIB	Libellé du quartier d'immatriculation
QAM_COD	Code du quartier d'immatriculation
QAM_RG	Rang géographique du quartier
SEG_SIHECO_LIB	Libellé du segment spécifique d'appartenance
SEG_SIHECO_COD	Code du segment spécifique d'appartenance
F_SIHECO_LIB	Libellé de la flottille spécifique d'appartenance
F_SIHECO_COD	Code de la flottille spécifique d'appartenance
SEG_DCR_LIB	Libellé du segment DCR
SEG_DCR_COD	Code du segment DCR
F_DCR_LIB	Libellé de la flottille DCR
F_DCR_COD	Code de la flottille DCR
SEG_DCR_Standard	Segment DCR standard
CARN_EFFECTIF	Effectif moyen à bord
NB_MOIS_ACT	Nombre de mois d'activité
FILEMO	Produit des valeurs des 3 composantes NAVP_LONGUEUR_HT, CARN_EFFECTIF et NB_MOIS_ACT (voir partie 2.2)
ind_Sortie_N_1	Indicateur binaire de sortie de flotte au cours de l'année N-1
ind_Present3112_N_2	Indicateur binaire de présence en flotte au 31 décembre de l'année N-2
ind_ChgtArm_N_1	Indicateur binaire de changement d'armateur au cours de l'année N-1
ind_ChgtFac_N_1	Indicateur binaire de changement de façade au cours de l'année N-1
ind_Inactifs	Indicateur binaire d'inactivité au cours de l'année N-2
ind_BaseSondage	Indicateur binaire d'inclusion du navire dans la base de sondage (déduit des 5 indicateurs précédents)

2.1.5 En résumé

Une base de sondage annuelle pour la collecte des données économiques en métropole (collecte opérée au cours de l'année N sur l'activité de l'année N-1) est constituée pour chacune des deux façades maritimes sur la base des navires présents dans le FPC de l'année de référence N-1. Les informations individuelles délivrées par le FPC sont enrichies par une typologie des navires fournie par l'Ifremer et mise en place à partir des calendriers d'activité portant sur l'année N-2 (source la plus actualisée disponible à cet instant). Cet ensemble d'informations individuelles permet ainsi de caractériser les critères d'inclusion des navires dans la base de sondage, de proposer les éléments qui composeront les stratifications utilisées ainsi que les supports pour les calculs d'allocation (FILEMO). Le décalage temporel des typologies disponibles constituera toutefois un biais modéré sur deux plans : la manière dont les navires seront affectés à une segmentation, et par extension, leur potentialité à être enquêtés puisque la validité de l'indice d'inactivité, qu'il soit positif ou négatif, sera empreint d'une relative incertitude.

2.2 Calculs des paramètres d'ajustement du chiffre d'affaires annuel individuel par la variable composite FILEMO

2.2.1 Présentation de la méthode

La méthode d'allocation d'effort de collecte stratifiée mise en place s'appuie sur des **objectifs de précision visés sur la stratification** considérée. Cet ajustement d'effectifs enquêtés vers un coefficient de variation (CV) d'estimateurs maîtrisé doit s'appliquer sur une **variable individuelle exhaustive au niveau de la population ciblée**, et de plus suffisamment corrélée aux indicateurs clés qui seront évalués à partir des données échantillonnées, afin que leur précision se rapproche autant que possible de la valeur ciblée initialement. La variable exhaustive choisie, et mise à contribution dans ce cadre d'application, sera un **proxy du chiffre d'affaires annuel**, image d'une fonction puissance appliquée à une variable composite particulière (fonction résultant donc d'une régression linéaire log-log). On appellera ce proxy « **variable auxiliaire** », ou encore « **variable d'optimisation** ».

Cette variable composite, appelée **FILEMO**, est le **produit de 3 variables individuelles** issues des données FPC et des données des calendriers d'activité : NAVP_LONGUEUR_HT (longueur individuelle du navire, en cm), CARN_EFFECTIF (effectif moyen annuel embarqué à bord du navire) et NB_MOIS_ACT (nombre de mois d'activité du navire⁸) (Van Iseghem, S. et *al.*, 2011). Elle est intégrée chaque année au sein de la base de sondage, et sert à la constitution d'un **modèle prédictif du chiffre d'affaires individuel** qui fournira le **proxy exhaustif**. Ce modèle est ajusté tous les ans sur les données de chiffre d'affaires issues des enquêtes opérées durant les années N-2, N-3 et N-4 (on le rappelle, pour une collecte de données sur N-1 opérée durant l'année N). Deux versions de ce modèle sont implémentées, une pour chaque façade.

Pour cela, sur les données disponibles pour les trois années considérées, on procède à un **ajustement linéaire log-log** (ajustement linéaire du logarithme du CA par le logarithme de FILEMO), générant un couple (a,b) d'intercept et de pente, tel que $a+b \cdot \log(\text{FILEMO})$ soit un proxy de $\log(\text{Chiffre d'Affaires})$. Le passage au logarithme permet de linéariser le lien initial de type « puissance » qui relie la variable Chiffre d'Affaires à FILEMO, simplifiant ainsi la problématique posée. L'utilisation de la régression linéaire présente également l'avantage de faciliter le processus de validation de l'ajustement obtenu. Un rapport d'ajustement est envoyé chaque

⁸ On considère que le navire a été actif à la pêche un mois donné à partir de moment où il a été actif à la pêche au moins 1 jour au cours de ce mois.

année au SSP, décrivant en particulier les deux couples de valeurs estimées (un pour la façade AT, l'autre pour la façade ME), afin que la variable support du calcul d'allocation soit évaluée au niveau individuel sur l'ensemble de la population, et que l'algorithme de détermination des allocations d'effort de collecte puisse ainsi être lancé.

2.2.2 Synthèse de l'analyse des résultats obtenus : application AT

On décrit ici la procédure d'ajustement de FILEMO appliquée à la façade Manche-Mer du Nord-Atlantique. Les étapes demeurent les mêmes pour l'application sur la façade Méditerranée.

On commence par explorer le lien entre $\log(\text{FILEMO})$ et $\log(\text{CA})$, en distinguant l'année considérée, au moyen d'un simple nuage de points. **Tester l'effet « année »** permet en premier lieu de valider le choix d'ajustement sur une période de 3 années par la **stabilité de la relation** qui doit être établie entre les deux variables.

Malgré une dissémination plus prononcée des chiffres d'affaires (CA) correspondant aux plus petites valeurs de FILEMO, dénotant une corrélation fragilisée pour ces valeurs extrêmes, on distingue déjà une **nette tendance au lien linéaire sur la globalité des points**. **L'effet « année » ne semble pas ressortir, du moins graphiquement**. L'ajustement s'exerce sur la globalité des navires enquêtés sur la période allant de l'année N-5 jusqu'à l'année N-3, et aboutit à une **droite de régression**.

Le couple de valeurs ainsi déterminé sera utilisé pour calculer le proxy du CA à partir des valeurs individuelles FILEMO pour les navires appartenant à la façade Manche-Mer du Nord-Atlantique, à partir de la formule :

$$\text{CA} = e^a \times \text{FILEMO}^b$$

où a est l'intercept et b la pente.

Il ne faut pas négliger que le couple obtenu (a,b) est bien le couple définissant le lien linéaire optimal entre $\log(\text{FILEMO})$ et $\log(\text{CA})$, mais qu'une fois l'égalité passée à l'exponentielle (fonction non linéaire) pour établir le lien entre FILEMO et CA, cette optimalité n'est alors plus qu'approximative. Nous vérifions alors que la valeur de b trouvée ne s'écarte pas trop de la valeur de b telle que :

- le **coefficient de corrélation linéaire** entre CA et FILEMO^b est **maximal** (on rappelle que ce coefficient permet de mesurer la liaison *linéaire* qui existe entre deux variables quantitatives)
- le **coefficient de variation** de FILEMO^b est **similaire** à celui de CA (on rappelle que le CV d'une variable numérique est une mesure relative de la dispersion des données autour de la moyenne).

Les deux graphiques ci-dessous (cf. figure 1) permettent de vérifier la validité de la valeur de b retenue.

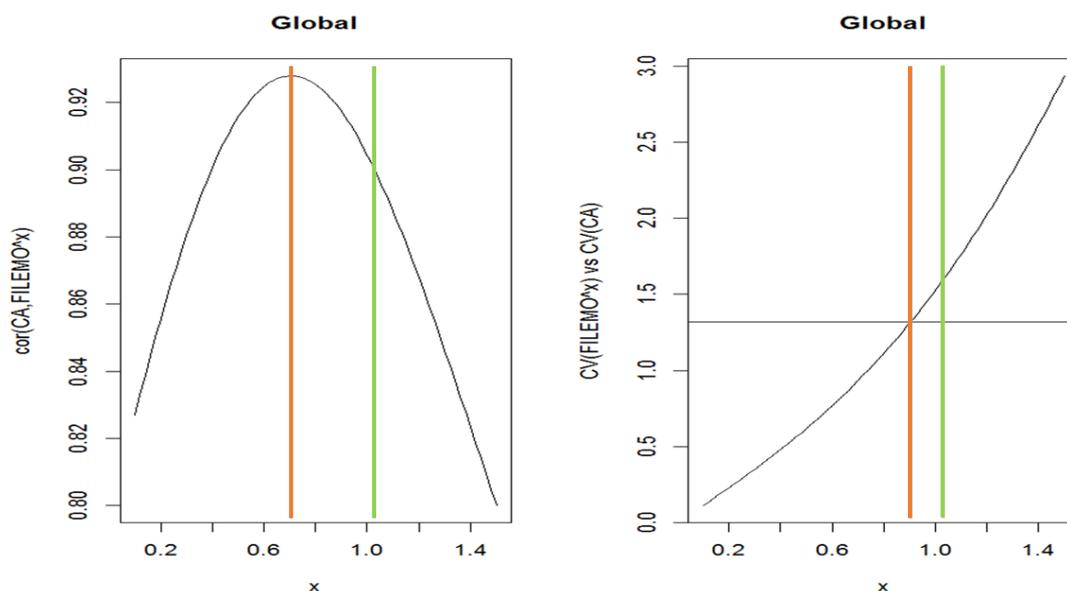


Figure 1 : Exemples de graphes de validation des paramètres retenus (à gauche, en termes de corrélation linéaire, à droite, en termes de variation). Pour chacun, en orange, le paramètre « puissance optimal », en vert le paramètre b ajusté. (Sources : Données administratives du fichier "Flotte de Pêche communautaire" (FPC), Données Activité des navires -Ifremer - Système d'Informations Halieutiques (SIH), SSP - Enquête sur la production de données économiques dans le secteur des pêches maritimes).

2.3 Méthodologie d'allocation appliquée au plan d'échantillonnage : problème d'optimisation de Neyman sous contraintes

La **méthode d'allocation** utilisée dans le cadre de la construction du plan d'échantillonnage pour la collecte des données économiques est basée sur **une optimisation de Neyman** (minimisation de la variance de l'estimateur moyen global) **sous diverses contraintes**. Elle vise à **déterminer** selon une stratification donnée **les effectifs à enquêter** en fonction de besoins particuliers, également appelées contraintes, quant à la **qualité statistique** des indicateurs issus de la collecte.

La contrainte prioritaire dans notre cas sera une contrainte de **précision locale**, qui pourra s'appliquer au sein de la segmentation initiale caractérisant le plan d'échantillonnage stratifié (indexée h dans ce qui suit), ou bien encore au niveau agrégé d'un regroupement de ces segments (indexé p). Cette précision se rapporte le plus souvent à un **coefficient de variation** (CV_p), mais peut tout aussi bien s'exprimer en tant qu'erreur relative maximale avec une probabilité donnée.

Une autre contrainte repose sur **l'encadrement des effectifs** alloués par segment/strate, ceux-ci étant compris entre un effectif minimal fixé, et un effectif maximal qui est naturellement la population du segment considéré. Enfin, une troisième contrainte, secondaire dans le cadre de notre plan d'échantillonnage car non limitante sur les objectifs fixés, est une **allocation totale fixe**. Cette dernière découle le plus souvent d'une contrainte de type **budgétaire**. On peut donc formaliser ce problème de la façon suivante :

$$\left\{ \begin{array}{l} \text{Min}_{n_1, \dots, n_H} V(\hat{Y}) = \sum_h \left(\frac{N_h}{N} \right)^2 \cdot \left(1 - \frac{n_h}{N_h} \right) \cdot \frac{S_h^2}{n_h} \quad (1) \\ n_{h \min} \leq n_h \leq N_h, \forall h = 1, \dots, H \quad (2) \\ \sum_{h=1, \dots, H} n_h = n \quad (3) \end{array} \right.$$

où h désigne le segment, n_h l'allocation correspondante, N_h la population totale correspondante, Y la variable exhaustive considérée (donc ici notre variable d'optimisation, le proxy du chiffre d'affaires issu de FILEMO ; voir chapitre précédent), et S_h^2 la variance de la variable Y au sein du segment. La valeur $n_{h \min}$ est l'allocation par segment h permettant d'atteindre l'objectif de précision fixé CV_p , avec p désignant le regroupement de segments sur lequel porte la précision visée (si nos objectifs de précision portent sur la stratification initiale, on peut poser $p=h$). Elle incarne donc bien ici l'allocation minimale selon les contraintes de précision à atteindre, et conditionne la contrainte (2) d'encadrement de l'allocation. Il faut noter dans ces conditions qu'une allocation supérieure à $n_{h \min}$ équivaut à une valeur de CV résultant inférieure à CV_p , c'est-à-dire à une précision améliorée au sein du regroupement (Koubi et al, 2009). La valeur n (nombre total d'observations) conditionne quant à elle la contrainte (3) d'allocation totale. Enfin, la contrainte (1) représente l'objectif d'optimisation dans le cadre de (2) et de (3).

2.4 Tirage de l'échantillon et répartition entre partenaires de collecte

Une fois la variable exhaustive issue de FILEMO construite comme approximation du chiffre d'affaires individuel des navires, et une fois celle-ci utilisée comme variable d'optimisation dans la procédure de calcul des allocations stratifiées, l'étape suivante va consister à tirer l'échantillon selon les effectifs ainsi distribués. On procède pour cela à un **tirage aléatoire systématique stratifié par segment d'allocation** (défini comme la combinaison « segment DCF » et « opérateur de collecte »). Celui-ci consiste, pour chaque strate d'appartenance à (cf. Figure 2) :

- isoler dans la base de sondage les navires appartenant à la strate (par exemple, tous les chalutiers et senneurs de fond de moins de 10m visés par l'enquête Ifremer) ;
- rapprocher le nombre total de navires de la population dans cette strate N à l'allocation correspondante n ;
- **ordonner** les navires **selon le quartier maritime d'appartenance** (répartition côtière, du nord au sud pour la façade AT, d'ouest en est pour la façade ME) **puis selon la longueur** quantitative du navire (en cm) ;
- **séparer** l'ensemble des navires **en n sous-ensembles d'effectifs similaires** $\left(\left\lfloor \frac{N}{n} \right\rfloor \right)$ individus par sous-ensemble, avec un individu supplémentaire pour $N - n \times \left\lfloor \frac{N}{n} \right\rfloor$ sous-ensembles répartis de manière homogène sur l'ensemble). Par exemple, si $N=46$ et $n=7$, alors les effectifs des sous-ensembles seront $E1=7, E2=6, E3=7, E4=6, E5=7, E6=6$ et $E7=7$ (voir figure 2) ;
- **tirer aléatoirement de manière équiprobable** un entier i entre 1 et $\left\lfloor \frac{N}{n} \right\rfloor$, et sélectionner le i -ème élément de chaque sous-ensemble. Les individus ainsi sélectionnés constitueront l'échantillon pour la strate considérée (une autre

possibilité consiste à tirer un entier i entre 1 et $\lfloor \frac{N}{n} \rfloor + 1$, et de prendre le i -ème élément du sous-ensemble lorsque cela est possible, le $(i-1)$ -ème sinon).

Navire	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20	N21	N22	N23	N24	N25	N26	N27	N28	N29	N30	N31	N32	N33
Quartier	Q1	Q2	Q2	Q2	Q2	Q2	Q3	Q4	Q4	Q4	Q5	Q5	Q5	Q5	Q5																		
Taille	573	607	692	706	788	919	960	625	772	773	779	795	519	543	549	609	615	622	643	726	765	785	804	935	955	508	686	713	516	628	655	745	784

Figure 2 : Mode de découpage de la base de sondage en sous-ensembles avant le tirage systématique

Une fois le tirage aléatoire effectué sur chacune des strates, un dernier arbitrage est réalisé afin de valider (ou non) l'assignation des navires à chaque opérateur de collecte, aboutissant finalement à la répartition définitive des navires à enquêter entre les deux partenaires.

La collecte est réalisée par le Lemna et l'Ifremer selon deux procédures distinctes et complémentaires. Le Lemna collecte les données à partir des comptabilités d'entreprise. L'Ifremer collecte directement les données auprès des patrons/armateurs des navires de pêche par voie d'enquêtes réalisées par des enquêteurs répartis sur le littoral. Quand l'échantillon est tiré, le Lemna indique pour quels navires il doit pouvoir collecter l'information grâce à la mobilisation de son réseau⁹. L'échantillon est ensuite réparti entre le Lemna et l'Ifremer.

Il découle de ce mode de collecte une répartition très différenciée selon la taille. En effet, les plus gros navires tiennent tous une comptabilité et ont beaucoup plus de chance d'être approchés par le Lemna (voir figure 3). En Méditerranée, où les navires de petite taille sont beaucoup plus nombreux, la part de navires enquêtée par l'Ifremer est sensiblement plus importante que sur la façade Atlantique.

⁹ Leur réseau de partenaires situés sur tout le littoral français, constitués de Centres de gestion agréés, de Cabinets d'expertise comptable ou d'Armements à la pêche.

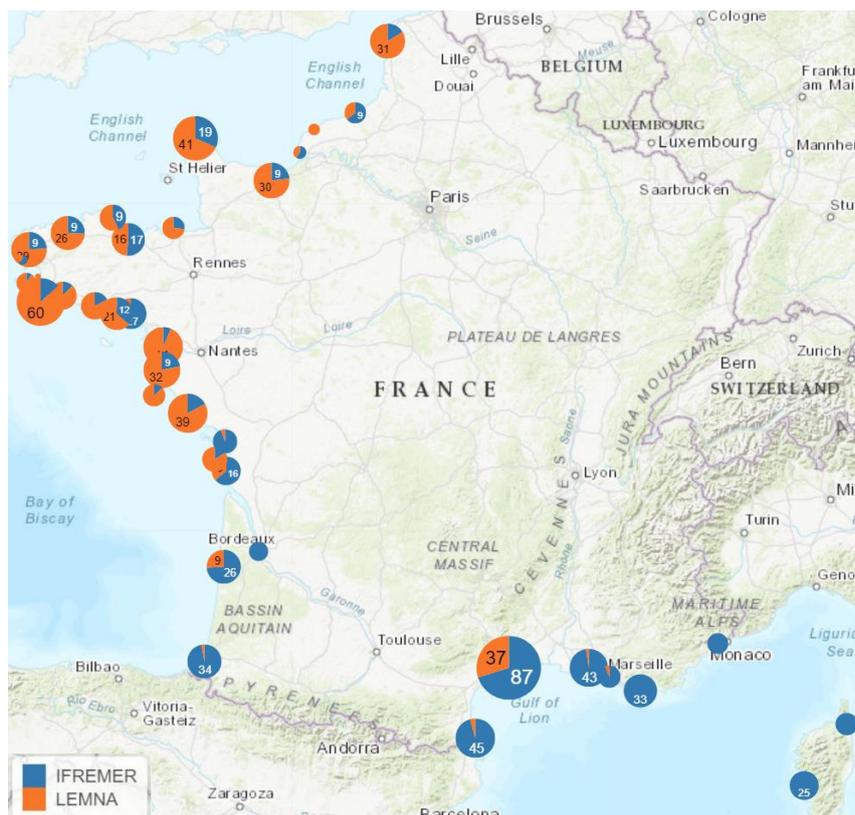


Figure 3 : Répartition des navires 'collectés' selon le quartier maritime et l'organisme enquêteur (Lemna en orange, Ifremer en bleu) en 2017. (Source : Gitton et al., 2020)

Une fois la répartition des navires tirés dans le plan d'échantillonnage national entre les deux partenaires de collecte, le travail d'organisation budgétaire et logistique de la campagne de collecte peut alors commencer pour chacun des partenaires, en fonction des navires leur ayant été alloués.

3 Références bibliographiques

Gitton François-Pierre, Minne Marie-Dominique, Baranger Laurent, Souffez Arnaud, Guyader Olivier, Le Grand Christelle, Merzereaud Mathieu (2020). Enquête pour la production de données économiques dans le secteur des pêches maritimes 2021-2025. Dossier de présentation au Comité du label de la statistique publique. Séance du 21 octobre 2020. 288p. <https://archimer.ifremer.fr/doc/00694/80622/>

Koubi, Malik; Mathern, Sandrine Résolution d'une des limites de l'allocation de Neyman, Actes des Xèmes Journées de Méthodologie Statistique, Volume 2009 (2009). http://www.jms-insee.fr/2009/S22_3_ACTE_KOUBI_JMS2009.PDF

Van Iseghem, S., Quilléro, E., Brigaudeau, C., Macher, C., Guyader, O., and Daurès, F. (2011). Ensuring representative economic data: survey data-collection methods in France for implementing the Common Fisheries Policy. – ICES Journal of Marine Science, 68: 1792–17