

Mise à jour de programmes de calcul de capture par unité d'effort pour les sous-populations de hareng du sud du Golfe du Saint-Laurent par la traduction du langage SAS vers le langage R

Lysandre Landry, François Turcotte, Nicolas Rolland, Jenni L. McDermid, Aurélie Cosandey-Godin et Andrew B. Harbicht

Pêches et Océans Canada
Centre des Pêches du Golfe
343 Avenue de l'Université
Moncton, Nouveau-Brunswick
Canada E1C 9B6

2024

**Rapport manuscrit canadien des
sciences halieutiques et aquatiques 3286**



Rapport manuscrit canadien des sciences halieutiques et aquatiques

Les rapports manuscrits contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles, mais qui traitent de problèmes nationaux ou régionaux. La distribution en est limitée aux organismes et aux personnes de régions particulières du Canada. Il n'y a aucune restriction quant au sujet; de fait, la série reflète la vaste gamme des intérêts et des politiques de Pêches et Océans Canada, c'est-à-dire les sciences halieutiques et aquatiques.

Les rapports manuscrits peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports manuscrits sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports manuscrits sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page du titre.

Les numéros 1 à 900 de cette série ont été publiés à titre de Manuscrits (série biologique) de l'Office de biologie du Canada, et après le changement de la désignation de cet organisme par décret du Parlement, en 1937, ont été classés comme Manuscrits (série biologique) de l'Office des recherches sur les pêcheries du Canada. Les numéros 901 à 1425 ont été publiés à titre de Rapports manuscrits de l'Office des recherches sur les pêcheries du Canada. Les numéros 1426 à 1550 sont parus à titre de Rapports manuscrits du Service des pêches et de la mer, ministère des Pêches et de l'Environnement. Le nom actuel de la série a été établi lors de la parution du numéro 1551.

Canadian Manuscript Report of Fisheries and Aquatic Sciences

Manuscript reports contain scientific and technical information that contributes to existing knowledge but which deals with national or regional problems. Distribution is restricted to institutions or individuals located in particular regions of Canada. However, no restriction is placed on subject matter, and the series reflects the broad interests and policies of Fisheries and Oceans Canada, namely, fisheries and aquatic sciences.

Manuscript reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts*.

Manuscript reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Numbers 1-900 in this series were issued as Manuscript Reports (Biological Series) of the Biological Board of Canada, and subsequent to 1937 when the name of the Board was changed by Act of Parliament, as Manuscript Reports (Biological Series) of the Fisheries Research Board of Canada. Numbers 1426 - 1550 were issued as Department of Fisheries and Environment, Fisheries and Marine Service Manuscript Reports. The current series name was changed with report number 1551.

Rapport manuscrit canadien des
sciences halieutiques et aquatiques 3286

2024

MISE À JOUR DE PROGRAMMES DE CALCUL DE CAPTURE PAR UNITÉ D'EFFORT
POUR
LES SOUS-POPULATIONS DE HARENG DU SUD DU GOLFE DU SAINT-LAURENT PAR
LA TRADUCTION DU LANGAGE SAS VERS LE LANGAGE R

Lysandre Landry¹, François Turcotte¹, Nicolas Rolland¹, Jenni L. McDermid¹, Aurélie Cosandey-Godin² et Andrew B. Harbicht¹

¹ Centre des Pêches du Golfe, Pêches et Océans Canada, 343 Avenue de l'Université
Moncton, New Brunswick, E1C 9B6, Canada

² Transport Canada, 700 Place Leigh-Capreol
Dorval, Québec, H4Y 1G7, Canada

© Sa Majesté le Roi du chef du Canada,
représenté par le ministre du ministère des Pêches et des Océans, 2024
N° de cat. Fs97-4/3286F-PDF ISBN 978-0-660-71516-2 ISSN 1488-5468

On doit citer la publication comme suit :

Landry, L., Turcotte, F., Rolland, N. , McDermid, J. L., Cosandey-Godin, A. et Harbicht, A. B. 2024. Mise à jour de programmes de calcul de capture par unité d'effort pour les sous-populations de hareng du sud du Golfe du Saint-Laurent par la traduction du langage SAS vers le langage R. Rapp. manus. can. sci. halieut. aquat. 3286 : x + 66 p.

RÉSUMÉ

Landry, L., Turcotte, F., Rolland, N. , McDermid, J. L., Cosandey-Godin, A. et Harbicht, A. B. 2024. Mise à jour de programmes de calcul de capture par unité d'effort pour les sous-populations de hareng du sud du Golfe du Saint-Laurent par la traduction du langage SAS vers le langage R. Rapp. manus. can. sci. halieut. aquat. 3286 : x + 66 p.

La pêche aux harengs du sud du golfe du Saint-Laurent des divisions 4T et 4Vn de l'Organisation des pêches de l'Atlantique Nord-Ouest est composée de deux sous-populations nommées : reproducteurs d'automne et de printemps. Elle est effectuée par la flotte côtière aux filets maillants et par la flotte de bateaux à senne coulissante. La pêche aux filets maillants se déroule sur les frayères à l'automne, alors qu'au printemps il s'agit d'une pêche d'interception à proximité des frayères. L'évaluation des deux sous-populations repose sur des modèles qui utilisent l'estimation de la capture par unité d'effort (CPUE) comme l'une des sources de données initiales. Les données utilisées pour l'estimation sont obtenues par trois différents relevés effectués à l'aide de la flotte côtière aux filets maillants seulement. Pour chacune des sous-populations, un modèle multiplicatif est utilisé pour obtenir une estimation standardisée de la CPUE par année. Ce travail avait pour but de mettre à jour les scripts utilisés pour l'estimation de la CPUE par la traduction de ceux-ci du langage SAS vers le langage R. L'objectif de ce rapport est de décrire la procédure de calcul utilisée par le programme R et de présenter les différences entre les programmes R et SAS. Ce travail a permis d'améliorer la compréhension des données, la procédure de calcul utilisée, l'automatisation et la reproductibilité des résultats.

ABSTRACT

Landry, L., Turcotte, F., Rolland, N. , McDermid, J. L., Cosandey-Godin, A. et Harbicht, A. B. 2024. Mise à jour de programmes de calcul de capture par unité d'effort pour les sous-populations de hareng du sud du Golfe du Saint-Laurent par la traduction du langage SAS vers le langage R. Rapp. manus. can. sci. halieut. aquat. 3286 : x + 66 p.

The herring fishery of the southern Gulf of St. Lawrence in Northwest Atlantic Fisheries Organization Divisions 4T and 4Vn is composed of two subpopulations named: fall and spring spawners. It is carried out by a coastal gillnet fleet and by a purse seine fleet. Gillnet fishing takes place on the spawning grounds in fall, while in spring, it is an interception fishery close to the spawning grounds. The subpopulations assessment is based on models that use estimated catch per unit effort (CPUE) as one of the initial data sources. The data used for the estimation are obtained from three different surveys on the gillnet coastal fleet only. For each subpopulation, a multiplicative model is used to obtain a standardized CPUE estimate by year. The aim of this work was to update the scripts used for CPUE estimation by translating them from the SAS language into the R language. The objective of this report is to describe the calculation procedure used in the R program and to present the differences between R and SAS programs. This work has allowed to improve our understanding of the data, the calculation procedures, the automation and the result reproducibility of the results.

TABLE DES MATIÈRES

RÉSUMÉ.....	iii
ABSTRACT	iii
TABLE DES MATIÈRES	iv
LISTE DES TABLEAUX	vii
LISTE DES FIGURES	vii
INTRODUCTION.....	1
Estimation de la capture par unité d'effort (CPUE)	2
Traduction de SAS vers R.....	3
MÉTHODOLOGIE	3
Lancement manuel de l'exécution automatique des scripts	3
<i>0-run_cpue.R</i>	5
Les outils.....	6
Fonctions : A1-cpue_fct.R	6
<i>fix_net_to_fathoms_fct(x, avg)</i>	6
<i>std_name_fct(input_data)</i>	6
<i>std_date_fct(input_data)</i>	7
<i>yday_week_fct(input_data, language)</i>	7
Aires et agrégations spatiales du hareng du sGSL : <i>A2-herring_areas.R</i>	7
Comparaison des différentes agrégations spatiales	8
Sources de données	10
DMP : Programme de monitoring à quai	11
1a-dmp_load_fct.R	12
1b-dmp_load.R.....	12
1c-dmp_raw_mod.R	12
1d-dmp.R.....	13
Telsurvey : Relevé téléphonique.....	13
2a-telsurvey_load.R.....	14
2b-telsurvey.R	14
ZIF : fichier informatisé inter-zone	15
3a-zif_load.R	15
3b-zif.R.....	16
Calculs	16
Effort.....	16

Capture.....	17
CPUE.....	18
6a-cpue_data.R.....	18
6bi-wt_spring_reg.R.....	19
6b-cpue_model.R.....	19
RÉSULTAT.....	20
Lancement manuel de l'exécution automatique des scripts.....	20
Les outils.....	20
fix_net_to_fathoms_fct(x, avg).....	20
Utilité de la fonction fix_net_to_fathoms_fct.....	21
Aires spatiales du hareng du sGSL : <i>A2-herring_areas.R</i>	23
Sources de données.....	23
DMP : Programme de monitoring à quai.....	24
Telsurvey : Relevé téléphonique.....	27
ZIF : fichier informatisé inter-zone.....	29
Calculs.....	32
Effort.....	32
Capture.....	36
CPUE.....	37
DISCUSSION.....	43
Agrégations spatiales et aires de pêches.....	43
Mise à jour des scripts de données sources.....	44
DMP.....	44
Telsurvey.....	44
ZIF.....	44
La valeur du nombre de filets standards.....	45
Sélection des données de captures pour le printemps.....	46
Révision du calcul de la CPUE : le nombre d'observation.....	46
Conclusion.....	46
REMERCIEMENTS.....	47
RÉFÉRENCES.....	47
APPENDICE 1.....	50
APPENDICE 2.....	53
APPENDICE 3.....	54
Est-ce que les résultats sont similaires entre le DMP et le Telsurvey pour le nombre moyen de filets?.....	54

Serait-il préférable de combiner les 2 sources de données (DMP et Telsurvey)?	59
APPENDICE 4	60
Représentation du décalage entre les captures pour le DMP et les ZIF.....	60

LISTE DES TABLEAUX

Tableau 1. Table de correspondances des différentes agrégations et aires spatiales.	10
Tableau 2. Nombre d'observations par classe de taille évaluées par la fonction qui vérifie si la longueur d'un filet est en brasse <i>fix_net_to_fathoms_fct</i> pour le programme de monitorages à quai.	21
Tableau 3. Nombre d'observations par classe de taille évaluées par la fonction qui vérifie si la longueur d'un filet est en brasse <i>fix_net_to_fathoms_fct</i> pour le relevé téléphonique.	22

LISTE DES FIGURES

Figure 1. Aires spatiales utilisées dans les évaluations des sous-populations : Sous-divisions de l'OPANO (4Tf à 4Tq et 4Vn) où les couleurs correspondent aux régions élaborées selon le savoir traditionnel sur la distribution spatiale des aires de fraie (bleu : Nord, vert : Centre, rouge : Sud et fuchsia : 4Vn); et Aires de pêche ou zone de gestion de la pêche au hareng (16A à 16G, 17 et 18).	2
Figure 2. Diagramme de flux des opérations d'importation, de transformation et d'entreposage des données effectuées dans les scripts R. Le flux pour le calcul de la capture par unité d'effort (CPUE) est présenté en B (Figure 3).	4
Figure 3. Diagramme de flux des opérations effectuées par les scripts R nécessaires à la réalisation du calcul de capture par unité d'effort (CPUE) et l'entreposage des estimés. B réfère à la Figure 2.	4
Figure 4. Différentes agrégations spatiales utilisées dans l'évaluation des sous-populations de harengs du sud du Golfe. Le titre de chacun des panneaux réfère à l'agrégation illustrée. Chacun des cercles correspond à un numéro de communauté portuaire (<i>community_code</i>) formée de la concaténation des variables <i>province</i> , <i>district</i> , et <i>locality</i> et la couleur change en fonction de la modalité de l'agrégation. Les aires spatiales utilisées pour la gestion des pêches du hareng (16A-G, 17 et 18, voir Figure 1) sont illustrées à titre de référence.	9
Figure 5. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche printanière, obtenus par le DMP entre 1999 à 2021. R en bleu. SAS en orange.	25
Figure 6. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche automnale, obtenus par le DMP entre 1999 à 2021. R en bleu. SAS en orange.	26
Figure 7. Capture en tonne et nombre de voyages pour la pêche printanière, enregistrés par le DMP entre 1990 à 2021. R en bleu. SAS en orange.	27

Figure 8. Nombre moyen de filets standards <i>avg_nets</i> (boîte à moustaches) pour la pêche printanière par aire du relevé téléphonique <i>telsurvey_area</i> entre 1990 et 2021. R en bleu. SAS en orange.	28
Figure 9. Nombre moyen de filets standards <i>avg_nets</i> (boîte à moustaches) pour la pêche automnale par aire du relevé téléphonique <i>telsurvey_area</i> entre 1990 et 2021. R en bleu. SAS en orange.	29
Figure 10. Capture en tonne et le nombre de voyages pour la pêche printanière enregistrés par les ZIF entre 1990 et 2021 (hachuré), Pslips (pointillé) et la combinaison des ZIF et des Pslips (trait complet). R en bleu. SAS en orange.	31
Figure 11. Capture en tonne et le nombre de voyages pour la pêche automnale enregistrés par les ZIF entre 1990 et 2021 (hachuré), Pslips (pointillé) et la combinaison des ZIF et des Pslips (trait complet). R en bleu. SAS en orange.	32
Figure 12. Nombre moyen de filets standards utilisés pour la pêche printanière par aires du relevé téléphonique (1 à 8) entre 1990 et 2021. R en bleu. SAS en orange.	34
Figure 13. Nombre moyen de filets standards utilisés pour la pêche automnale par aires du relevé téléphonique (1 à 8) entre 1986 et 2021. R en bleu. SAS en orange.	35
Figure 14. Nombre moyen de filets standards mouillés ou relevés pour la pêche automnale par aires du relevé téléphonique (1 à 8) entre 1986 et 2021. R en bleu. SAS en orange.	36
Figure 15. Capture en tonne et du nombre de voyages pour la pêche printanière enregistrés par le DMP et complétés par les ZIF entre 1990 et 2021. R en bleu. R avec procédure SAS de sélection des ZIF en bleu pointillé. SAS en orange.	37
Figure 16. Capture en tonne et du nombre de voyages pour la pêche printanière enregistrés par le DMP et complétés par les ZIF entre 1990 et 2021. R en bleu. SAS en orange. Procédures avec (trait hachuré) et sans (trait plein) le retrait des <i>unknown_area</i> non applicable et égale à 9.	38
Figure 17. Capture en tonne et du nombre de voyages pour la pêche automnale enregistrés par les ZIF entre 1986 et 2021. R en bleu. SAS en orange. Procédures avec (trait hachuré) et sans (trait plein) le retrait des <i>large_area</i> non applicable et égale à 4Vn.	39
Figure 18. CPUE calculées (boîte à moustaches) par aire de pêche au hareng divisé par province (<i>unknown_area</i>), par semaine et par jour julien, pour la pêche printanière entre 1990 et 2021. R en bleu. SAS en orange.	40
Figure 19. Estimations de la CPUE (kg par levée de filet, moyenne et intervalles de confiance 5 et 95 %) pour la pêche printanière entre 1990 et 2021. R en bleu. SAS en orange.	40
Figure 20. CPUE calculées (boîte à moustaches) par région (<i>large_area</i> : Nord, Centre, Sud) et par semaine pour la pêche automnale entre 1986 et 2021. R en bleu. SAS en orange.	41

Figure 21. Estimations de la CPUE (kg par levée de filet, moyenne et intervalles de confiance 5 et 95 %) par région (<i>large_area</i> : Nord, Centre, Sud) pour la pêche automnale entre 1986 à 2021. R en bleu. SAS en orange.	42
Figure 22. Diagramme du flux des opérations illustrant les scripts SAS nécessaires à l'importation et la transformation et à l'archivage des données pour la réalisation du calcul de capture par unité d'effort (CPUE). La procédure illustrée par type de données sources : A : DMP (seulement les données du Golfe), B : DMP, C : ZIF, D : Telsurvey, est celle qui est généralement utilisée. Certaines différences sont observées entre les années.....	51
Figure 23. Diagramme du flux des opérations illustrant les scripts SAS nécessaires à l'importation des données archivées et à la réalisation du calcul de capture par unité d'effort (CPUE). La procédure illustrée : A : calcul de l'effort, B : estimation CPUE pour l'automne, C : estimation CPUE pour le printemps, est celle qui est généralement utilisée. Certaines différences sont observées entre les années.	52
Figure 24. Districts statistiques de l'Île-du-Prince-Edward (P.E.I.), du Nouveau-Brunswick (N.B) et de la Nouvelle-Écosse (N.S.), correspondant à la Figure 3 de Claytor et al. 1991.	53
Figure 25. Aire du relevé téléphonique 1 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	54
Figure 26. Aire du relevé téléphonique 2 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	55
Figure 27. Aire du relevé téléphonique 3 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	55
Figure 28. Aire du relevé téléphonique 4 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	56
Figure 29. Aire du relevé téléphonique 5 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	56
Figure 30. Aire du relevé téléphonique 6 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	57
Figure 31. Aire du relevé téléphonique 7 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	57

Figure 32. Aire du relevé téléphonique 8 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.	58
Figure 33. Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale pour toutes les aires du relevé téléphonique (1 à 8). DMP et gris et rouge. Telsurvey en blanc et bleu.	58
Figure 34. Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale pour toutes les aires du relevé téléphonique (1 à 8). DMP et gris et rouge. Telsurvey en blanc et bleu. Combinaison des deux en noir et vert.	59
Figure 35. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 1990 et 1999. DMP en gris. ZIF en noir hachuré.	61
Figure 36. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 2000 et 2009. DMP en gris. ZIF en noir hachuré.	62
Figure 37. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 2010 et 2021. DMP en gris. ZIF en noir hachuré.	63
Figure 38. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche automnale entre 2000 et 2009. DMP en gris. ZIF en noir hachuré.	64
Figure 39. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche automnale entre 2010 et 2021. DMP en gris. ZIF en noir hachuré.	65
Figure 40 Comparaison des nombres de districts inclus pour les enregistrements du DMP (pointillé), des ZIF (hachuré) et de la combinaison des 2 (trait plein) de 1990 à 2021 pour la pêche printanière et automnale.	66

INTRODUCTION

La population de hareng du sud du golfe du Saint-Laurent (sGSL) se divise en deux sous-populations génétiques distinctes nommées : les reproducteurs de printemps et les reproducteurs d'automne (Chadwick et Nielsen 1986; Lamichhane et al. 2017). Elles se mélangent sur les aires d'alimentation et d'hivernage, mais la reproduction est séparée temporellement et spatialement (Stobo 1982; Wheeler et al. 2006). La période de reproduction de la sous-population de printemps se déroule principalement en avril et en mai et celle d'automne de la mi-août à la mi-octobre. Bien qu'historiquement on sépare les périodes de reproduction par le 1^{er} juillet, il est possible d'observer des reproducteurs de printemps et d'automne sur les mêmes frayères au mois de juin (LeBlanc et al. 2007).

La pêche au hareng du sGSL couvre les divisions 4T et 4Vn (Chaput et al. 1993) de l'Organisation des Pêches de l'Atlantique Nord-Ouest (OPANO). Elle est subdivisée en quatre régions définies par le savoir traditionnel sur la distribution spatiale des aires de fraie (Sud, Centre, Nord et 4Vn; Chadwick et Cairns 1988; McDermid et al. 2016), ainsi que sur les aires de gestion du hareng (16A à 16G, 17 et 18; Chadwick et al. 1989; Figure 1). Elle est effectuée par la flotte côtière aux filets maillants (engins fixes; qui comprend aussi les petits senneurs de moins de 19,8 m), et par la flotte de bateaux à senne coulissante (engin mobile). La pêche aux filets maillants se déroule sur les frayères à l'automne, alors qu'au printemps il s'agit d'une pêche d'interception à proximité des frayères. Historiquement, au printemps, la plupart des filets sont ancrés pour la nuit et relevés le lendemain matin, tandis qu'à l'automne les filets sont installés lorsque des bancs de taille suffisante sont trouvés (Clayton et al. 1998). Ainsi, la pêche automnale aux filets maillants capture principalement des reproducteurs d'automne et la printanière des reproducteurs de printemps (Rolland et al. 2022). La pêche à la senne coulissante vise les géniteurs des deux sous-populations avant et pendant leurs migrations hors du sGSL au printemps et à l'automne (LeBlanc et al. 2012).

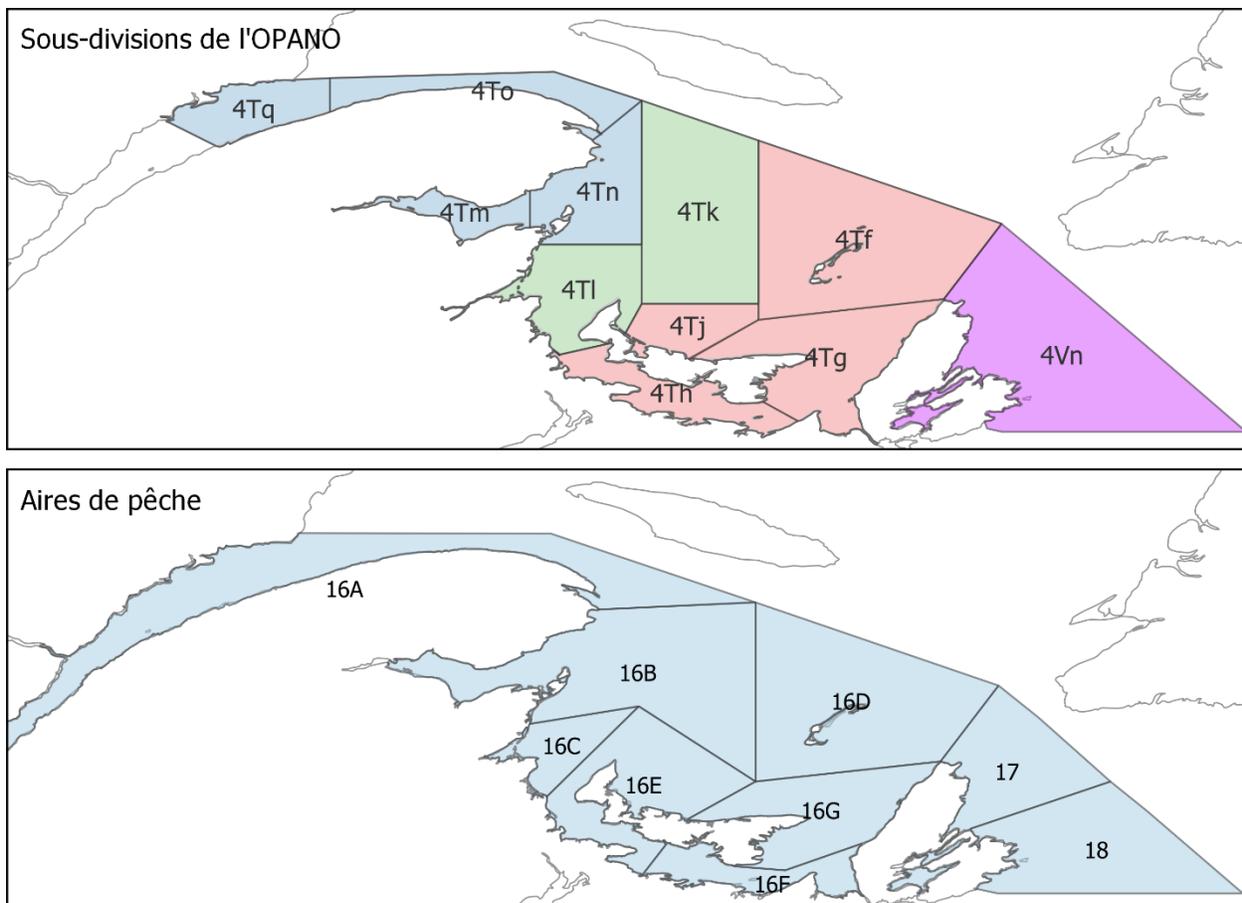


Figure 1. Aires spatiales utilisées dans les évaluations des sous-populations : Sous-divisions de l'OPANO (4Tf à 4Tq et 4Vn) où les couleurs correspondent aux régions élaborées selon le savoir traditionnel sur la distribution spatiale des aires de fraie (bleu : Nord, vert : Centre, rouge : Sud et fuchsia : 4Vn); et Aires de pêche ou zone de gestion de la pêche au hareng (16A à 16G, 17 et 18).

ESTIMATION DE LA CAPTURE PAR UNITÉ D'EFFORT (CPUE)

L'évaluation des deux sous-populations de reproducteur de hareng du sGSL utilise un modèle de population qui inclut un indice de capture à l'âge pour des individus de 4 à 11 ans. Cet indice est basé en partie sur l'estimation de la capture par unité d'effort (CPUE) de la pêche aux filets maillant (LeBlanc et al. 2009).

La CPUE est définie comme la biomasse au débarquement (capture) en kilogramme (kg) par levée de filet par voyage (effort) pour les reproducteurs d'automne et par la capture en kg par filet par voyage (effort) pour les reproducteurs de printemps (LeBlanc et al. 2009). Depuis l'évaluation de 1995 (Claytor et al. 1996), deux modèles multiplicatifs distincts : un pour le printemps et un pour l'automne, sont utilisés pour obtenir une estimation standardisée de la CPUE par année. Ils ont été actualisés au cours des années pour tenir compte des nouvelles connaissances et de la disponibilité des données et les modifications sont décrites dans les différents documents de recherche publiés.

TRADUCTION DE SAS VERS R

Ce travail avait pour objectif de mettre à jour les scripts utilisés pour l'estimation de la CPUE par la traduction de ceux-ci du langage SAS (*Statistical Analysis Software*; SAS Institute Inc. 2013) vers le langage R (*R Foundation for Statistical Computing Platform*; R Core Team 2021) et avait pour but d'améliorer l'accessibilité, la compréhension et la transférabilité des scripts. Un résumé de ce travail a été présenté lors de la revue par les pairs de l'évaluation des stocks de 2021 (Rolland et al. 2022).

Le premier objectif de ce rapport est de décrire la procédure utilisée par le programme R. Le second est de décrire les modifications qui ont été apportées à la procédure qui pourrait avoir un impact sur les résultats obtenus à la suite de l'exécution du programme R et le troisième est de comparer ces résultats avec ceux obtenus par le programme SAS.

MÉTHODOLOGIE

Le programme R a été élaboré à l'aide de la version R 4.1.2. (R Core Team 2021). Les scripts et documents requis sont regroupés dans un projet nommé *herring_cpue.Rproj* et archivés sur GC/Code; un service partagé du ministère des Pêches et Océans du Canada (MPO) et hébergé par les services GitLab.

LANCEMENT MANUEL DE L'EXÉCUTION AUTOMATIQUE DES SCRIPTS

Les scripts R sont regroupés en six sections de un à trois scripts où chacune des sections permet d'exécuter une tâche particulière pour toutes les années et les deux saisons simultanément. Un script supplémentaire, qui doit être modifié annuellement, appelle l'exécution de tous les autres : *0-cpue_run.R*. Ce dernier appelle aussi les scripts appelés Outils : *A1-cpue_fct.R* et *A2-herring_area.R*. Le flux d'exécution des scripts est représenté aux Figure 2 et Figure 3 et plus en détails dans les différentes sections de la méthodologie.

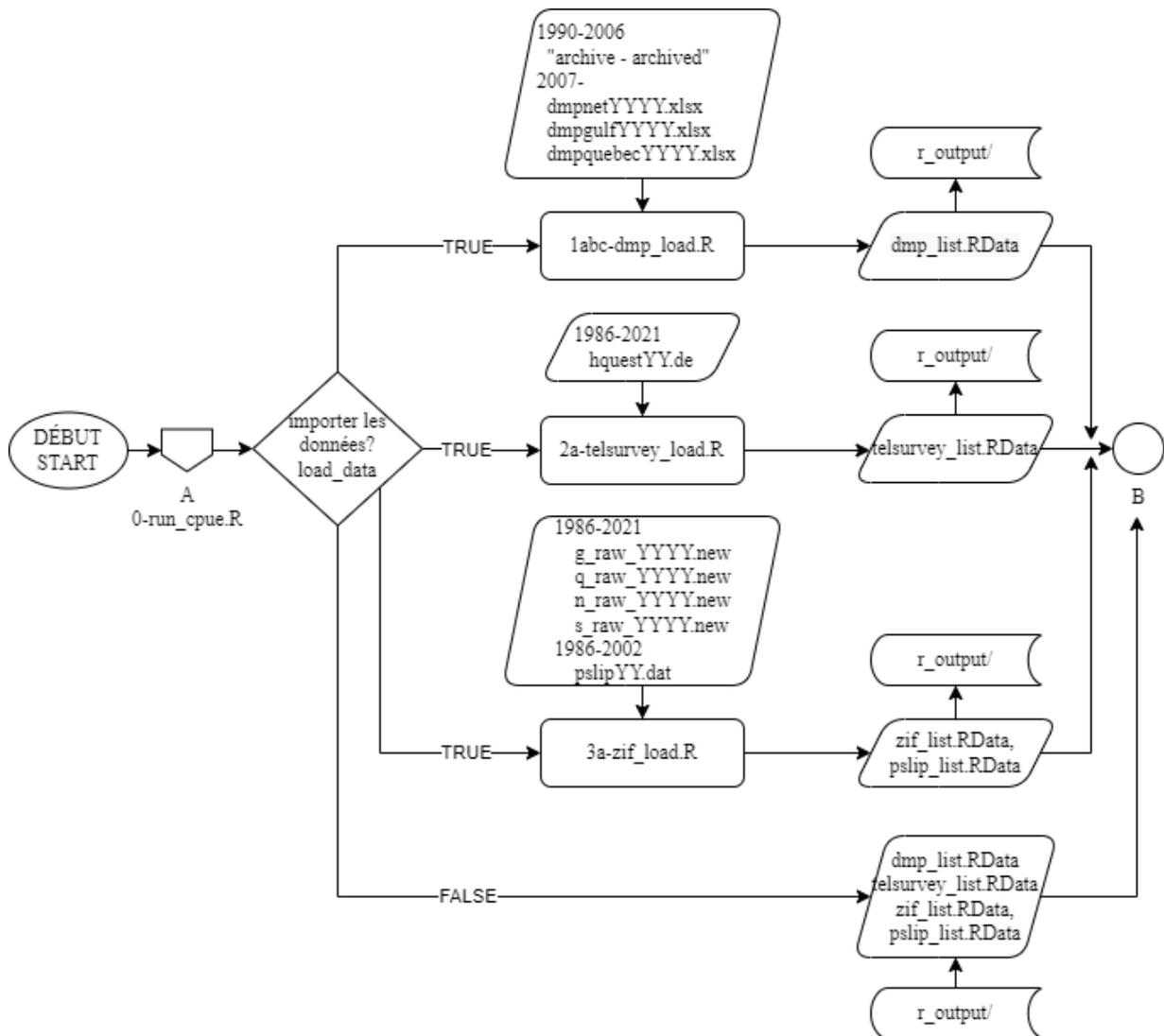


Figure 2. Diagramme de flux des opérations d'importation, de transformation et d'entreposage des données effectuées dans les scripts R. Le flux pour le calcul de la capture par unité d'effort (CPUE) est présenté en B (Figure 3).

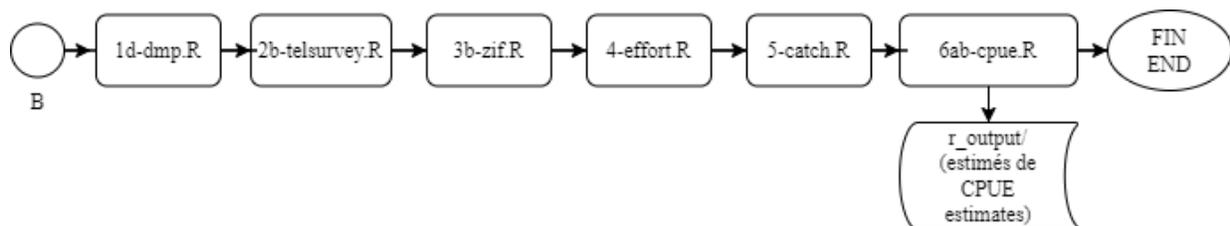


Figure 3. Diagramme de flux des opérations effectuées par les scripts R nécessaires à la réalisation du calcul de capture par unité d'effort (CPUE) et l'entreposage des estimés. B réfère à la Figure 2.

0-run_cpue.R

L'objectif de ce script est d'exécuter l'ensemble des scripts. De plus, il définit les répertoires où se trouvent les données, appelle les bibliothèques et définit les variables partagées par l'ensemble des scripts.

Lors de la première exécution, les adresses des différents répertoires où sont entreposées les données entrantes et sortantes doivent être modifiées. Généralement, ce sont les lettres assignées aux lecteurs *Hd2*, *Hd3* et *HerringM*, qui doivent être modifiées pour pouvoir avoir accès aux données. Ces lettres établissent la connexion avec les dossiers sur les sous-populations de hareng du sGSL entreposé sur le réseau du MPO. L'accès à la base de données Oracle (Oracle 2017) *PTRAN.glf_herring* est aussi nécessaire. Les répertoires d'accès vers les dossiers où se trouvent les documents requis sont aussi assignés à des variables globales qui seront utilisées ultérieurement :

- *fp_main*, entrepose l'adresse du répertoire du projet *herring_cpue.Rproj*.
 - Ainsi, les différents scripts R sont contenus dans le dossier *r_script/* du projet;
- *herring_m_dmp*, *hd2_spring* et *hd2_fall*, entreposent les adresses des répertoires où se trouvent les données du programme de monitoring à quai (DMP). Pour les années plus grandes que 2006, les fichiers de données sont aussi temporairement contenus dans le dossier : *std_dmp_data/* du projet;
- *hd2_phone*, entrepose l'adresse du répertoire où se trouvent les données de relevés téléphoniques (Telsurvey);
- *hd3_ziff* et *hd3_pslips*, entreposent les adresses des répertoires où se trouvent respectivement les données des fichiers informatisés inter-zones (ZIFF) et des bordereaux d'achat (Pslips), et;
- *herringm_sharbour* et *herringm_portqc*, entreposent les adresses des répertoires où se trouve l'information sur les ports de débarquement.

Les bibliothèques et paquets nécessaires aux fonctionnements des scripts sont : *data.table* (Dowle et Srinivasan 2021), *dplyr* (Wickham et al. 2021a), *emmeans* (Lenth 2022), *ggfortify* (Horikoshi et Tang 2016), *ggplot2* (Wickham 2016), *glue* (Hester et Bryan 2021), *gulf* (Surette et al. 2021), *kableExtra* (Zhu 2021), *librarian* (Quintans 2021), *lubridate* (Grolemund et Wickham 2011), *magick* (Ooms 2021), *readr* (Wickham et al. 2021b), *readxl* (Wickham et Bryan 2019), *rgdal* (Bivand et al. 2021), *RODBC* (Ripley et Lapsley 2021), *sp* (Pebesma et Bivand 2005), *stringr* (Wickham 2019) et *tidyr* (Wickham 2021).

Annuellement, pour exécuter ce script, une valeur doit être assignée aux variables suivantes :

- *yr_global*, la valeur numérique de l'année à évaluer;
- *load_data*, la valeur logique TRUE pour permettre d'importer les données de l'année en cours ou FALSE pour utiliser les données déjà importées et archivées dans le répertoire du projet et dossier *r_output/*. Cette variable permet de réduire le temps d'exécution du calcul lorsqu'elle est assignée à FALSE (voir Figure 2), et;

- *wt_spring_regression*, la valeur logique *TRUE* ou *FALSE* pour permettre ou non d'attribuer une pondération aux captures dans le modèle de régression qui produit les estimations standardisées la CPUE pour les reproducteurs de printemps.

Ensuite, l'utilisateur peut lancer l'exécution du script 0 qui entraîne l'exécution des scripts outils *A1-cpue_fct.R* et *A2-herring_areas.R* et la chaîne d'exécution de tous les scripts par l'appel du script *6b-cpue_model.R* (voir Figure 3). Les résultats obtenus à la suite de l'exécution sont sauvegardés dans le dossier *r_output/*.

LES OUTILS

Fonctions : *A1-cpue_fct.R*

L'objectif du script *A1-cpue_fct.R* est de créer des fonctions partagées par plusieurs scripts SAS, pour uniformiser les calculs et les variables utilisées. Aucun intrant n'est nécessaire à l'exécution de ce script. Il en résulte 4 fonctions.

fix_net_to_fathoms_fct(x, avg)

La fonction *fix_net_to_fathoms_fct* a comme attributs la valeur à valider (*x*) et la valeur de la longueur moyenne d'un filet en brasses (*avg*). Cette fonction permet de corriger les erreurs d'unité de mesure dans les champs de la variable *net_length_fm*, soit la longueur des filets en brasses pour les fichiers de données de débarquements du programme de monitoring des débarquements a quai (DMP pour *Dockside monitoring program*) et du relevé téléphonique (Telsurvey pour *Telephone survey*). Elle évalue si la valeur est en brasses et la convertit si elle est dans une autre unité de mesure, soit : verge ou pied, invalide ou non applicable. Avant 2014 inclusivement, la valeur moyenne (*avg*), dite standard, attribuée aux données invalides et non applicables était de 15 brasses. Depuis 2014, elle est de 14 brasses (McDermid et al. 2016).

Lors de l'exécution de la fonction :

- les valeurs plus petites que 10, plus grandes que 120 ou non applicables (NA) sont changées pour la valeur standard de l'année à laquelle les valeurs appartiennent;
- les valeurs entre 10 et 19 sont considérées comme étant en brasses (*fathom; fm*) et gardent leur valeur initiale;
- les valeurs entre 20 et 40 sont considérées comme étant en verge (*yard; yd*) et sont divisées par 2;
- les valeurs entre 60 et 120 sont considérées comme étant en pied (*foot; ft*) et sont divisé par 6, et;
- pour maintenir la cohérence avec les rapports précédents, les valeurs entre 50 et 56 sont assignées à la valeur 16,6 brasses.

std_name_fct(input_data)

La fonction *std_name_fct* a comme attribut *input_data*, soit le fichier de données sources ZIFF, Pslips, DMP ou Telsurvey à corriger. Cette fonction permet d'uniformiser et définir le nom d'une variable et sa classe (p. ex., caractère ou numérique). Elle retire

certaines variables qui ne sont pas requises ou intéressantes pour l'estimation de la CPUE.

std_date_fct(input_data)

La fonction *std_date_fct* a comme attribut *input_data*, soit le fichier de données sources ZIFF (pour *Zonal Interchange File Format*; Fichier de format d'échange intra-zone), Pslips (pour *Purchase slips*; bordereaux d'achats), DMP ou Telsurvey à corriger. Elle permet d'uniformiser le format des dates. Elle vérifie si les champs possèdent des valeurs aberrantes, par exemple :

- elle modifie un jour (*day*) égal à 31 pour les mois possédant 30 jours;
- retire les entrées où la date (*date*) ou le mois (*month*) n'a pas été saisie, mais conserve celles avec le jour (*day*) manquant, et;
- complète les champs où l'année (*year*) est manquante.

yday_week_fct(input_data, language)

La fonction *yday_week_fct* a comme attribut *input_data*, le fichier de données sources ZIFF, Pslips, DMP ou Telsurvey à modifier et; *language* qui permet de choisir entre la modalité SAS ou R, soit la procédure utilisée par le langage de programmation correspondant. Cette fonction permet d'ajouter la variable du jour julien (*yday*) ou celle du numéro de la semaine en fonction de l'année (*week*) au fichier de données.

Aires et agrégations spatiales du hareng du sGSL : *A2-herring_areas.R*

Les données des évaluations des sous-populations de harengs du sGSL sont entreposées et analysées selon différentes agrégations spatiales. Le script *A2-herring_areas.R* a pour objectif de créer un fichier de données qui rassemble toutes les agrégations spatiales utilisées, d'uniformiser les définitions, de définir le lien qui existe entre elles et de circonscrire la zone d'étude. Les tables de données Oracle :

- PTRAN.glf_herring.herring_wharf_code et
- PTRAN.glf_herring.new_herring_district,

ainsi que les fichiers :

- small_craft_harbours_coord.csv du répertoire herringm_scharbour, et
- port_qc.csv du répertoire herringm_portqc,

sont utilisées. Le fichier de données produit se nomme *areas.csv*.

Les numéros de communautés portuaires (*community_code*) identifient des ports où se produit des débarquements de pêches. Le premier chiffre des *community_code* correspond à la province (*province*), le deuxième et le troisième au district (*district*) et le quatrième et le cinquième à la localité (*locality*). Les variables enfants *homeport* (port d'attache) et *port_landed* (port de débarquement) de la variable *community_code* sont aussi utilisées dans l'ensemble des programmes de calcul. Les districts statistiques portuaires, *district* (Figure 24 de l'Appendice 2), se retrouvant dans une province, *province*, sont concaténés pour former la variable *prov_dist*. Cette nouvelle variable est

utilisée comme plus petit dénominateur commun pour définir les agrégations spatiales utilisées. Dans ce document, lorsqu'il sera question de districts statistiques ou de districts, cela correspondra à la variable *prov_dist*. Les agrégations spatiales de districts définies dans ce script sont représentées par les variables suivantes :

- *telsurvey_area*, est utilisée par le relevé téléphonique annuel et sont au nombre de formée de 9 modalités : 1 à 9;
- *herring_area*, est reliée aux zones de gestion des pêches ou à l'aire où s'effectue la pêche dans la région du sGSL et possède 9 modalités : 16A, 16B, 16C, 16D, 16E, 16F, 16G, 17 et 18;
- *unknown_area*, est utilisée pour les analyses qui correspondent aux *herring_area* séparées, de façon générale, par province. Elle peut être confondue avec la *telsurvey_area* puisqu'elle possède les mêmes 9 noms de modalité : 1 à 9, et enfin;
- *large_area*, correspond aux régions mentionnées dans les évaluations et possède 4 modalités : Nord (*North*), Centre (*Middle*), Sud (*South*), 4Vn.

Le script importe et modifie les deux tables Oracle et les fichiers .csv qui contiennent les informations relatives aux districts statistiques portuaires de la région du Golfe. Le format des données est modifié (*std_name_fct*) et les tables sont fusionnées. Les districts qui ne sont pas inclus dans la région du sGSL sont retirés. Les districts inclus correspondent à ceux inclus dans les aires de l'OPANO 4T et 4Vn, soit :

- pour la Nouvelle-Écosse (NS) ou province égale à 1 : les districts 01 à 04, 06, 07, 10 à 13, 45 et 46;
- pour le Nouveau-Brunswick (NB) ou province égale à 2 : les districts 63 à 68, 70, 71, 73, 75 à 78 et 80;
- pour l'Île-du-Prince-Édouard (PE) ou province égale à 3 : les districts 82, 83, 85 à 88, 92, 93, 95 et 96, et;
- pour le Québec (QC) ou province égale à 4 : les districts 01 à 18 et 26 à 28 (excluant la portion du Saint-Laurent à l'est de Pointe-des-Monts au Nord et à l'est la côte de la péninsule gaspésienne au sud).

Comparaison des différentes agrégations spatiales

La Figure 4 et le Tableau 1 permettent de comparer visuellement les différentes agrégations et aires spatiales utilisées dans les évaluations des sous-populations de hareng du sGSL.

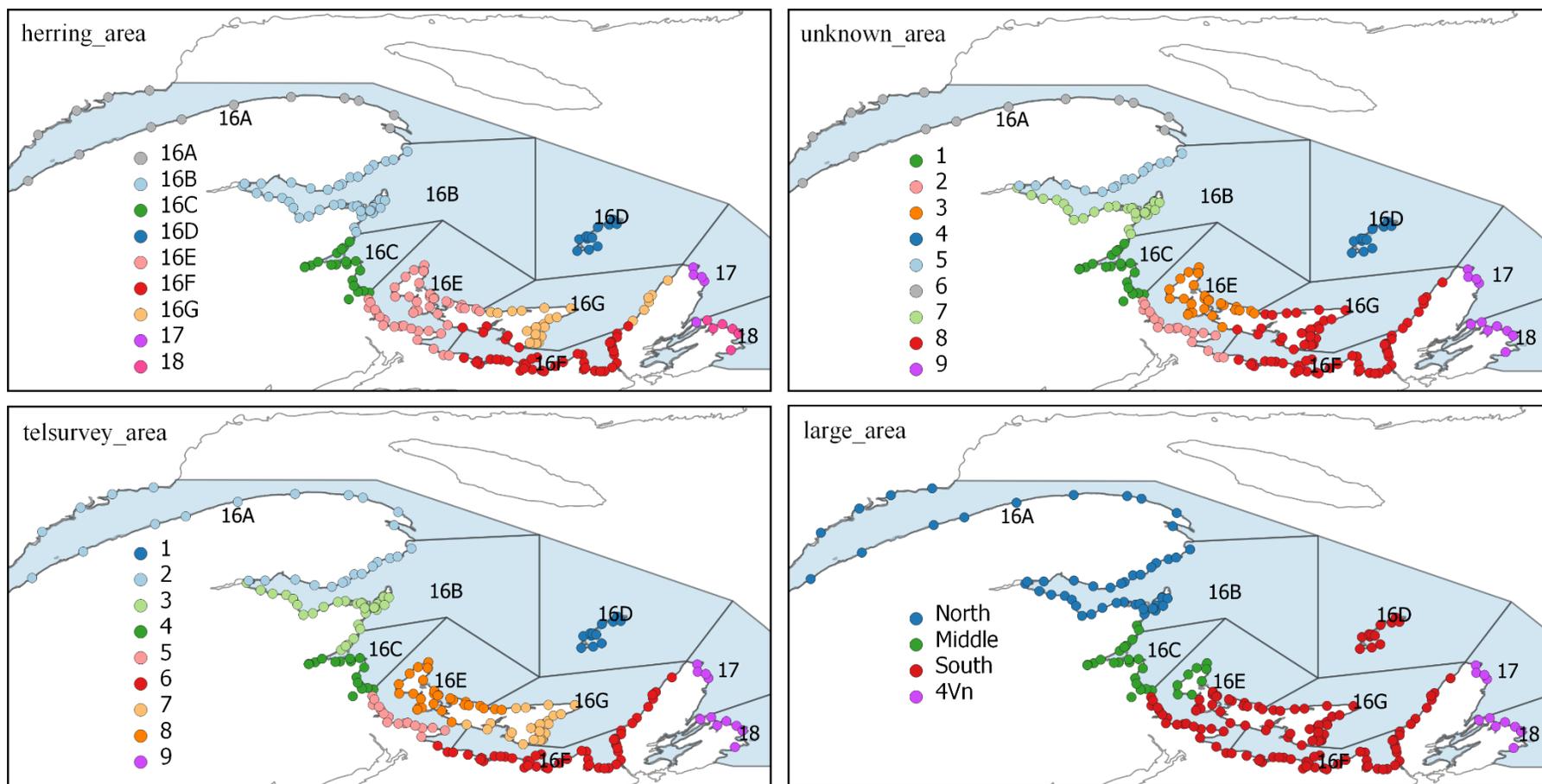


Figure 4. Différentes agrégations spatiales utilisées dans l'évaluation des sous-populations de harengs du sud du Golfe. Le titre de chacun des panneaux réfère à l'agrégation illustrée. Chacun des cercles correspond à un numéro de communauté portuaire (*community_code*) formée de la concaténation des variables *province*, *district*, et *locality* et la couleur change en fonction de la modalité de l'agrégation. Les aires spatiales utilisées pour la gestion des pêches du hareng (16A-G, 17 et 18, voir Figure 1) sont illustrées à titre de référence.

Tableau 1. Table de correspondances des différentes agrégations et aires spatiales.

<i>unknown_area</i>	<i>herring_area</i>	<i>telsurvey_area</i>	<i>large_area</i>	<i>province</i>	<i>district</i>
1	16C	3, 4	Middle	NB	70, 71, 73, 75, 76
2	16E	5, 6	South	NB, NS	45, 77, 78, 80
3	16E	8	Middle, South	PE	82, 83, 92, 93, 95
4	16D	1	South	QC	26 à 28
5	16B	2	North	QC	10 à 15
6	16A	2	North	QC	1 : 9, 16 à 18
7	16B	3	Middle, North	NB	63 : 68
8	16F, 16G	6, 7, 8	South	NS, PE	2, 3, 10 à 13, 46, 85 : 88, 96
9	17, 18	9	4Vn	NS	1, 4, 6, 7

SOURCES DE DONNÉES

L'estimation de la CPUE est basée sur trois relevés ou sources de données principales :

- le programme de monitoring des débarquements à quai (DMP) qui recueille quotidiennement des données sur la capture en kg par bateau depuis 1990 (Clayton et al. 1996) et sur le nombre et la longueur des filets en brasse utilisés par voyage depuis 1999;
- le relevé téléphonique annuel (Telsurvey) qui recueille des données sur différentes composantes de l'effort dont le nombre de filets, la longueur des filets en brasse et le nombre de levées de filet, par saison et par bateau (ou responsable du bateau) depuis 1986 (LeBlanc et LeBlanc 1996), et;
- le fichier informatisé inter-zone (ZIFF) qui collige et fournit l'information sur les débarquements quotidiens totaux déclarés (captures quotidiennes en kg) et recensés par le DMP et par les bordereaux d'achat (Pslips). Il est utilisé comme source de données principale ou complémentaire aux données du DMP et des Pslips selon les années et la saison de pêche. Les Pslips permettent d'obtenir des données sur les captures en kg vendues quotidiennement par bateau par communauté portuaire qui sont utilisées pour les années entre 1986 à 2002 inclusivement (Chadwick et Nielsen 1987).

Ainsi, pour chacune de ces sources, il y a une ou plusieurs : sources de données et/ou fichiers de données brutes et/ou fichiers de données transformées et archivées, selon les années. Pour chacun des fichiers, un premier script crée une fonction d'importation, un autre importe les données et un dernier modifie les données.

Pour le calcul de la CPUE, la plus petite unité statistique (ou unité d'échantillonnage) est un voyage de pêche. De façon générale, il y a qu'un seul voyage de pêche par jour, donc l'unité peut être décrite de deux façons, soit :

- un débarquement par numéro de bateau (*cfv*), par jour julien (*yday*) pour un numéro de port de débarquement donné (*port_landed*), ou;
- un débarquement par numéro de bateau (*cfv*), par voyage (*trip_id*), par jour julien (*yday*) pour un numéro de port de débarquement donné (*port_landed*).

Le DMP et les ZIFF conservent cette unité. Cependant, puisque le relevé téléphonique est effectué annuellement, l'unité statistique pour ce relevé est :

- numéro de bateau (*cfv*) ou responsable (*resp*), par saison (*season*), par aire du relevé (*telsurvey_area*).

Ainsi, un fichier de données est produit par source de données : DMP, Telsurvey et ZIFF. Chacun des fichiers de données rassemble toutes les années pour les deux saisons de pêches. Ils ont pour objectif d'uniformiser les données et de conserver la plus petite résolution de l'unité possible.

DMP : Programme de monitoring à quai

Les scripts dont la numérotation débute par 1 manipulent les données du programme de monitorages à quai. L'objectif de ces scripts est d'importer les données brutes de l'année en cours en provenance de la région du Golfe ou du Québec; de les transformer pour obtenir le nombre de filets standard (*nets_std*) pour le calcul de l'effort et; de les transformer pour obtenir les captures en kg (*catch_kg*) pour le calcul de la capture par année (*year*), par mois (*month*), par jour (*day*), par port de débarquement (*port_landed*), par bateau (*cfv*) et par voyage (*trip_id*).

Les données brutes n'existent plus pour les années 1990 à 2006 inclusivement, alors les données transformées par les programmes SAS archivées en MS Excel sont utilisées. À partir de 2007 jusqu'à *yr_global*, les données brutes provenant de la division des statistiques du MPO sont utilisées. Par contre, elles ont été archivées en MS Excel et les noms des variables, des feuilles de calculs et des fichiers ont été modifiés manuellement au cours des années. Étant donné la complexité de la structure et de l'organisation de ces données brutes archivées en MS Excel, les données ont dû être recollées dans des fichiers MS Excel plus uniformisés pour faciliter l'importation.

Les données se trouvent :

- dans le dossier *std_dmp_data/* (données uniformisées) ou le répertoire *herring_m_dmp* pour les données brutes de 2007 à *yr_global*, et;
- le répertoire *hd2_spring*, pour les données du DMP transformées par les scripts SAS de 1990 à 2006.

Les fichiers produits par R et nécessaires aux calculs subséquents se nomment *dmp_std* pour le calcul de l'effort et *dmp_catch* pour le calcul de la capture.

1a-dmp_load_fct.R

L'objectif du script *1a-dmp_load_fct.R* est de créer les fonctions nécessaires à l'importation des données en provenance du programme de monitoring à quai (DMP). Ce script produit les fonctions suivantes :

- *readdmp_oldsas_fct(ext)*, où l'attribut *ext* de la fonction correspond au nom du fichier à importer. La fonction a pour objectif d'importer les données 1999-2006 qui n'ont pas de données brutes associées;
- *readdmp_fct(tab, source, yr)*, où les attributs : *tab*, *source* et *yr*, de la fonction correspondent respectivement au nom de la feuille MS Excel, à la source des données et à l'année des observations incluses dans le fichier. La fonction a pour objectif d'importer les données brutes où l'attribut *tab* peut correspondre à 4 modalités soit : *spring* et *fall*, correspondant respectivement aux données du printemps et de l'automne; *bait*, correspondant aux données des appâts pour la région du Québec seulement, et; *all*, correspondant à la somme des autres feuilles MS Excel. Chaque fichier MS Excel est associé à une seule provenance (*source*) des données soit : pour la région du Québec; *dmpquebec*, pour la région du Golfe; *dmpgulf* et pour les données sur les filets de la région du Golfe; *dmpnet*.

Chacune des fonctions contient la variable globale définie dans le script 0 qui entrepose l'adresse du répertoire où se trouvent les données requises.

1b-dmp_load.R

L'objectif du script *1b-dmp_load.R* est d'importer les données du DMP à l'aide des fonctions créées précédemment.

Il uniformise les données à l'aide des fonctions *std_name_fct* et *std_date_fct* et les colle dans les listes suivantes :

- *dmp_old_list* inclut les données transformées pour les années 1990 à 2006 pour lesquels il n'existe plus de données brutes. Les éléments de la liste sont nommés d'après l'ancien nom du fichier soit : *nbcoord_1990_1998*, *magdmp.1999_2003*, *gulfdmp.1999_2003*, *gulf_que_dmpcorr.2004*, *gulf_que_dmp.2005*, *gulf_que_dmp.2006*;
- *dmp_raw_list* inclut les données brutes pour les années 2007 à *yr_global*. Les éléments de la liste sont nommés par année (ex. 2007).

Enfin, le programme sauvegarde les listes produites dans le dossier *r_output/*. Cette sauvegarde est utilisée pour ne pas avoir à importer les données plus d'une fois par année. Ainsi, le script sera utilisé seulement si *load_data == TRUE* (Figure 2) ou si la liste requise n'est pas présente dans l'environnement global du projet.

1c-dmp_raw_mod.R

L'objectif du script *1c-dmp_raw_mod.R* est d'effectuer des vérifications et corrections automatiques sur les données brutes contenues dans la liste *dmp_raw_list* créée précédemment.

Le script :

- supprime les lignes dupliquées; les lignes où il manque une information sur les variables *cfv*, *year*, *month*, *day* et *port_landed* et; les lignes avec des données aberrantes de la variable *cfv*;
- corrige pour les lignes où, pour une journée (*day*), une sortie d'un bateau (*cfv*) équivalent à un *trip_id*, il existe deux données de capture (*catch_kg*);
- effectue la somme des captures et la moyenne des tailles de maille de filet (*mesh_size_in*). Une moyenne est effectuée sur la taille de maille, car elle n'est présentement pas utilisée dans le calcul de l'effort.
- corrige pour certaines lignes la valeur de la variable *nets_fished* lorsque la valeur de *nets_set* est plus grande que la valeur de la variable *nets_fished*. Le script sélectionne la valeur la plus élevée des 2. La variable *nets_set* est définie comme le nombre de fois que le filet a été mouillé ou relevé. La variable *nets_fished* est définie comme le nombre de filets utilisés multiplié par le nombre de *nets_set* (McDermid, pers. comm. 2021).

1d-dmp.R

L'objectif du script *1d-dmp.R* est de sélectionner et de modifier les données nécessaires à l'obtention du nombre de filets standard (*nets_std*) pour la saison de printemps et d'automne et de la capture en kg par voyage (*catch_kg* par *trip_id*) pour le printemps seulement. Il utilise les listes *dmp_old_list* et *dmp_raw_list* et produit le fichier *dmp_std* pour le calcul de l'effort et *dmp_catch* pour le calcul de la capture.

Pour créer le fichier de données utilisé pour le calcul de l'effort; *dmp_std*, le script :

- standardise la longueur des filets en brasses, à l'aide de la fonction *fix_net_to_fathoms_fct*, selon les années auxquelles ils appartiennent (McDermid et al. 2016);
- retire les lignes qui sont associées à la capture pour les appâts, soit les lignes où le nombre de filets relevés est entre 1 et 6 au printemps;
- supprime les lignes où il n'y a pas d'information sur le nombre de filets relevés;
- crée un nombre standardisé de filets à partir de la longueur moyenne d'un filet établi à partir de l'évaluation de stock de 2014 (McDermid et al. 2016);
- sélectionne les variables essentielles au calcul effectué ultérieurement, et;
- retire les lignes dupliquées.

Pour le fichier utilisé pour le calcul de la capture : *dmp_catch*, les variables qui caractérisent les filets sont retirées et, à la suite de cette sélection, les lignes dupliquées sont à nouveau retirées.

Telsurvey : Relevé téléphonique

Les scripts dont la numérotation débute par 2 manipulent les données du relevé téléphonique annuel. L'objectif de ces scripts est d'importer les données brutes de l'année en cours; de les transformer pour obtenir le nombre de filets standard (*avgnets*)

pour le calcul de l'effort par année (*year*), par saison (*season*), par aire du relevé téléphonique (*telsurvey_area*), par bateau (*cfv*) et par responsable de bateau (*resp*).

Les données brutes se trouvent sur le répertoire *hd2_phone* et l'extension des fichiers est *.de*. Le fichier produit se nomme *telsurvey_std*.

2a-telsurvey_load.R

L'objectif du script *2a-telsurvey_load.R* est d'importer les données et de combiner les données des relevés téléphoniques de chaque année, d'effectuer les corrections.

Ainsi, il effectue les corrections annuelles, celles pour les erreurs de saisies et celles pour les erreurs de format de variable équivalent à la fonction *std_names_fct*. Enfin, il entrepose les données dans une liste de fichiers appelée *telsurvey_list.RData*.

2b-telsurvey.R

Le script *2b-telsurvey.R* importe la liste avec les données brutes corrigées et effectue la correction pour la longueur des filets à l'aide de la fonction *fix_net_to_fathoms_fct*.

Pour le relevé par téléphone, le nombre moyen de filets utilisés *avgnets* est déterminé à l'aide de trois variables :

- *peaknets*; le nombre de filets utilisés lors d'une saison d'abondance de capture au moment où il y a une abondance de capture;
- *restnets*; le nombre de filets utilisés lors d'une saison d'abondance de capture au moment où il y a une faible abondance de capture et;
- *npeaknets*; le nombre de filets utilisés lors d'une saison où il n'y a pas de période d'abondance de capture.

Comme il est fait pour les données du DMP, le script retire, pour les trois variables précédentes, les nombres de filets inférieurs ou égaux à 6 pour la saison de printemps, car ils sont associés aux captures qui servent d'appât; et, il calcule le nombre standardisé de filets à partir de la longueur moyenne d'un filet établi pour de l'évaluation de stock de 2014.

Le script fait ensuite le calcul d'un nombre moyen de filets standard *avgnets*.

- Le script vérifie si la pêche a lieu durant une saison où il y a un pic d'abondance ou non (*peak* égale à Y ou N). Certaines lignes du fichier de données n'ont pas d'information pour cette variable. Lorsque *peak* égale à NA et que la valeur de *peakdays* (le nombre de jours de pêche où on observe une forte abondance) est présente, *peak* est assignée à Y.
- Lorsque *peak* égale Y,
 - le script détermine combien de jours il y avait sans pic d'abondance soit : *npdays* plus grand que 0. Normalement, *npdays* égale à *totdays* (le nombre total de jours de pêche) moins *peakdays*. Cependant, il arrive que *totdays* ne soit pas comptabilisé et que *peakdays* le soit ou que *peakdays* soient plus élevés que *totdays*. Dans ces cas, le plus grand nombre est assigné à *totdays* pour ensuite déterminer la valeur de *npdays*.

Ensuite, le script détermine le nombre de filets utilisés pour les *npdays* plus grand que 0. Normalement, il devrait être inscrit à la variable *restnets*, mais si le nombre est absent, il assigne à *resnets* la quantité de *npeaknets*. Si la valeur de *resnets* est toujours absente et qu'il y a une valeur à *peaknets* le script lui assignera sa valeur.

Lorsque *npdays* égale à 0 et que *peaknets* n'a pas de valeur, mais que *restnets* en a une. Le script assigne la valeur de *restnets* à *peaknets*.

Enfin, l'assignation à *avgnets* se fait de la façon suivante :

- *avgnets* égale à *peaknets*
 - S'il existe des jours sans pic d'abondance *npdays* et que *peaknets* n'a pas de valeur *avgnets* égale à *restnets*
 - Si *restnets* et *peaknets* ont une valeur, c'est la valeur moyenne qui est assignée à *avgnets*.
- Pour les saisons sans pic d'abondance, où *peak* égale à N :
 - Le nombre de jours de pêche sans pic d'abondance *npdays* dans une saison sans grande abondance est égal à *totdays* le nombre total de jours de pêche.
 - Le nombre de filets standards utilisés dans ces jours sans pic *npnets* est *npeaknts* et donc *avgnets* est égale à *npeaknts*.

Le fichier produit s'appelle *telsurvey_std*.

ZIFF : fichier informatisé inter-zone

Les scripts dont la numérotation débute par 3 manipulent les données des ZIFF. Les données contenues dans les ZIFF ont pour objectif d'informer sur la capture en kg par année (*year*), par mois (*month*), par jour (*day*) ou voyage (*trip_id*) et par port de débarquement (*port_landed*).

3a-zif_load.R

L'objectif du script *3a-zif_load.R* est d'importer les données des ZIFF en provenance des fichiers de données brutes qui se trouvent dans les répertoires *hd3_zif* (ZIFF; 1986 à *yr_global*) et *hd3_pslip* (bordereaux d'achat; 1986 à 2002).

Pour les données des ZIFF et des Pslips, le script sélectionne les données du hareng (*species.code* égale à 200) de la zone 4T (*nafo.subdivision* 4Tg à 4Tq) pour les engins de pêche utilisant des filets maillant (*gear.code* 40 à 43). Particulièrement, il sélectionne les captures plus petites que 25 000 kg (20 000 kg pour les années inférieures ou égales à 2007), les numéros de bateau (*cfv*) entre 1 et 189999 inclusivement et les ports de débarquement (*port_landed*) de 10100 à 11399 et de 20000 à 49999. Cette sélection est ensuite entreposée par année dans les listes *zif_list* et *pslip_list* en fonction de la source à laquelle elle appartient.

3b-zif.R

L'objectif du script *3b-zif.R* est de sélectionner les données nécessaires à l'obtention de la capture en kg (*catch_kg*) par voyage (*trip_id*). Il utilise les listes précédentes, retire les doublons et les captures égales à zéro, puis sélectionne la capture la plus élevée entre les deux sources de données des ZIFF ou des Pslips par année (*year*), mois (*month*), jour (*day*), port de débarquement (*port_landed*) et numéro de bateau (*cfv*) pour produire le fichier *zif_pslip*.

CALCULS

Cette section rassemble les étapes relatives à la sélection des données, à l'agrégation et au calcul de l'effort, de la capture et de la CPUE.

Effort

L'objectif de cette section et du script *4-effort.R* est de calculer le nombre moyen de filets standardisés à 14 brasses (*nets_std_mean*) pour chaque aire de relevé téléphonique (*telsurvey_area*), saison (*season*) et année (*year*).

Depuis l'évaluation des stocks de 2007, le nombre moyen de filets standard est obtenu par deux sources de données soit le DMP et le Telsurvey, le script choisit les valeurs en provenance de la source de données qui possède le plus grand nombre d'observations, soit : le nombre de bateaux (*cfv*) et/ou de responsables de bateau (*resp*) (LeBlanc et al. 2008). De plus, depuis 2008, le calcul effectué est différent entre le printemps et l'automne puisqu'à l'automne le calcul de l'effort inclut le nombre de levées de filet (Claytor et al. 1998; LeBlanc et al. 2009), bien que pour le printemps le nombre de levées de filet est considéré comme étant de un par jour (Claytor et al. 1998). Ce choix s'effectue en trois étapes.

Premièrement, le nombre moyen de filets standardisés est obtenu par l'agrégation des données par année, saison, aire de relevé téléphonique, numéro de bateau et/ou responsable pour les données du relevé téléphonique et du programme de monitoring à quai. Une moyenne est effectuée pour les lignes où l'information sur le bateau et/ou le responsable de bateau est disponible, mais toutes les autres lignes où *cfv* ou *resp* est égal à NA sont conservées.

Pour effectuer la moyenne, la variable du nombre de filets standards (*net_std*) du fichier *dmp_std* est utilisée pour le DMP. Cette variable inclut le nombre de levées de filet (*hauls*) par définition de la variable *nets_fished* avec laquelle elle est élaborée. Pour le Telsurvey, c'est la variable du nombre moyen de filets standardisés *avgnets* qui est utilisée et, pour l'automne seulement, *eff2* si une information pour celle du nombre de levées de filet *hauls* existe en provenance du fichier *telsurvey_std*. La variable *eff2* est obtenue de la façon suivante :

$$eff2 \leftarrow ((peakdays * peaknets) + (npdays * npnets) * hauls) / totdays$$

Deuxièmement, un nombre d'observations est obtenu par le calcul du nombre de bateaux ou de responsables ayant fourni une information. Une moyenne est aussi calculée pour chacune des sources de données. Les variables *n* et *net_std_mean* pour

le DMP et les variables *n*, *net_std_mean* et *eff2_mean* pour le Telsurvey sont produites. Dans quatre fichiers de données :

- *Telsurvey_area* et *dmp_area* contiennent les données agrégées par année, saison et aire du relevé téléphonique, et;
- *Telsurvey_year* et *dmp_year* contiennent les données agrégées par année et saison.

Troisièmement, les fichiers *Telsurvey_area* et *dmp_area* sont fusionnés et par unité statistique: année, saison, aire du relevé téléphonique, le script sélectionne la source de données ayant le plus d'observation. Le fichier en résultant se nomme *dmp_telsurvey_choice*. La même procédure est effectuée pour l'agrégation par année et saison. Le fichier en résultant se nomme *dmp_telsurvey_choicey*.

Quatrièmement, pour obtenir les valeurs de l'effort soit : les nombres moyens de filets standardisés (*net_std_mean*) ou les nombres moyens de filets standardisés relevés (*eff2_mean*) par année, par saison de pêche et aires du relevé téléphonique, une procédure séquentielle est utilisée où la première valeur est écrasée par la suivante si elle existe.

1. Un fichier vide comprenant les variables année (*year* : 1986 à *yr_global*), saison (*season* : *spring*, *fall*), aires du relevé téléphonique (*telsurvey_area* : 1 à 8) est créé.
2. Les valeurs du fichier *dmp_telsurvey_choice* agrégées par saison, par année et par aire du relevé téléphonique sont ajoutées, puis;
3. les valeurs du fichier *dmp_telsurvey_choicey* agrégées par saison et par année sont ajoutées.

Le fichier *dmp_telsurvey_choice* contient les résultats de cette procédure.

Capture

L'objectif de cette section et du script *5-catch.R* est de calculer la capture en kg (*tot_catch_kg*) et le nombre de voyages (*trips*), par saison (*season*), par année (*year*), par mois (*month*), par jour (*day*), et par province et district (*prov_dist*).

Le calcul effectué est différent entre le printemps et l'automne, car en 1995, une analyse a montré qu'au printemps, il y avait un décalage entre les captures hebdomadaires enregistrées des ZIFF et des DMP (Claytor et al. 1995). Depuis, les valeurs de capture pour le printemps proviennent des données du DMP pour les districts où a lieu le monitoring et des ZIFF pour les autres districts. Pour l'automne, c'est la capture provenant des ZIFF qui est conservée.

La plus petite unité statistique pour les données du DMP et les ZIFF est une capture en kg (*catch_kg*), par journée ou par sortie de pêche (*day* ou *trip_id * day*), par bateau (*cfv*), pour un numéro de port de débarquement (*port_landed*). Ce script calcule la somme des captures (*tot_catch_kg*) et le nombre de voyages (*trips*), par saison (*season*), par année (*year*), mois (*month*), jour (*day*), district statistique (*prov_dist*). Le *prov_dist* provient de la variable *port_landed*.

Pour les données du printemps seulement, le script prend les données des ZIFF provenant des districts (*prov_dist*) qui ne sont pas présents dans les données du DMP

pour toute la série temporelle (1990 à *yr_global*) tels que décrit dans les rapports de recherches reliés aux évaluations de stock de hareng du sGSL (Rolland et al. 2022).

Le fichier *catch_all* est produit.

CPUE

L'objectif de cette section est d'estimer la capture par filet standard (*ce*) par année (*year*), pour la pêche du printemps et la capture par levée de filet standard (*ce*) par année (*year*), pour chacune des régions (*large_area*), pour celle d'automne.

6a-cpue_data.R

L'objectif du script *6a-cpue_data.R* est de sélectionner et d'agrèger les données nécessaires pour le calcul de la capture par unité d'effort (*ce* ou *cpue*; logarithme naturel de *ce*) pour le printemps (kg par filet) et l'automne (kg par levée de filet), par année (*year*), mois (*week*), jour julien (*yday*), et district (*prov_dist*).

L'exécution du script ajoute les variables du mois (*month*) et du jour julien (*yday*) au fichier *catch_all* à l'aide de la fonction *yday_week_fct* où l'attribut *language* est égale à SAS. Puis, les variables de l'aire du relevé téléphonique (*telsurvey_area*), des régions (*large_area*) et des aires du hareng divisées par province (*unknown_area*) sont ajoutées au fichier *catch_all* à l'aide d'une jointure par *prov_dist* effectuée à l'aide du fichier *areas* qui établit le lien entre les différentes agrégations spatiales.

Ensuite, le script effectue la fusion par saison (*season*), année (*year*) et aire du relevé téléphonique (*telsurvey_area*), du fichier d'effort *dmp_telsurvey_choice* qui contient les variables *nets_std_mean* et *eff2_mean* avec; le fichier de capture *catch_all* qui contient les variables de la capture en kg (*tot_catch_kg*) et du nombre de voyages (*trips*).

Pour obtenir la variable *effort*, le script multiplie le nombre de voyages (*trips*) avec le nombre moyen de filets standards (*nets_std_mean*). Pour obtenir la variable *effort2*, le script multiplie le nombre de voyages (*trips*), avec le nombre de filets relevés (*eff2_mean*), s'il existe, ou dans le cas contraire, il utilise la variable *effort*. Pour la saison du printemps, la variable *effort* est toujours utilisée et assignée à *effort2*.

Le script sélectionne ensuite les données utilisées dans l'estimation soit :

- Pour le printemps : les semaines (*week*) 12 à 22 (mi-mars à fin-mai) et les années (*year*) 1990 à *yr_global* et,
- Pour l'automne : les semaines (*week*) 27 à 43 (début juillet à fin octobre) et les années (*year*) 1986 à *yr_global*.

Le mois de juin est retiré pour éviter le mélange des géniteurs d'automne et de printemps dans les captures (LeBlanc et al. 2007). La justification pour le choix de l'année du début de la série temporelle égale à 1990 pour le printemps et à 1986 pour l'automne se trouve dans les documents suivants : Chaput et al. 1993; Claytor et al. 1995, 1996 et; LeBlanc et al. 2009.

Pour le printemps, le script crée un fichier de données : *cpue_tel_yday*, qui fait la somme sur les variables *effort2* et *tot_catch_kg*, par *season*, *year*, *unknown_area*, *week* et *yday*. Pour l'automne, il crée un fichier de données : *cpue_week*, qui fait la somme

par *season*, *year*, *large_area* et *week*. La justification de ces choix d'agrégation pour les modèles se trouve dans les documents suivants : LeBlanc et al. 2002; Poirier 2002 et McDermid et al. 2016.

La CPUE (*ce*) est ensuite calculée en effectuant la division de la capture (*tot_catch_kg*) par le nombre de levées ou le nombre de filets (*effort2*) pour chacun des fichiers.

Pour respecter la Loi sur la protection des renseignements personnels (L.R.C. (1985), ch. P-21; « *Privacy Act* ») qui spécifie que les pêcheurs ne puissent pas être identifiés lors de la présentation des résultats, toutes les lignes dont le nombre de voyages (*trips*) est inférieur ou égal à 5 sont retirées à cette étape du calcul.

6bi-wt_spring_reg.R

L'objectif du script *6bi-wt_spring_reg.R* est de calculer la proportion *wt* de la somme des captures de l'année *catch_year* appartenant à un jour julien soit :

$$wt \leftarrow catch_yday / catch_year$$

où *catch_yday* est la somme des captures d'un jour julien c'est-à-dire la capture par *yday*, *season*, *year* et *unknown_area*. Cette proportion sera utilisée pour l'estimation de la CPUE du printemps seulement.

6b-cpue_model.R

L'objectif du script *6b-cpue_model.R* est d'estimer la CPUE (moyenne des moindres carrés) à l'aide de deux modèles de régression élaborés de façon distincte pour chacune des sous-populations.

L'estimation de la CPUE est basé sur la formule suivante pour le printemps :

$$\ln CPUE_{ijk} = \alpha + \beta_1 I + \beta_2 J + \beta_3 K + \varepsilon \quad (1)$$

où *I* réfère à l'année, *J* réfère à l'aire de gestion du hareng divisée par province, *K* réfère à la semaine et ε est l'erreur résiduelle. Ce modèle utilise les données de CPUE (*ce*) par aire de gestion du hareng divisé par province (*unknown_area*), par année, par semaine et par jour julien, et; estime la CPUE par année. La CPUE dans le modèle est pondérée selon la proportion de la capture annuelle attribuée à un jour julien (*wt*; LeBlanc et al. 2002).

Pour l'automne, la formule utilisée est :

$$\ln CPUE_{ik} = \alpha + \beta_1 I + \beta_3 K + \varepsilon \quad (2)$$

où *I* réfère à l'année, *K* réfère à la semaine et ε est l'erreur résiduelle. Le modèle est utilisé pour trois des quatre régions (*large_area*): Nord, Centre et Sud séparément. Le modèle utilise les données de la CPUE (*ce*), par année, par semaine et; estime la CPUE par année (McDermid et al. 2016).

Le script R exécute le modèle suivant pour l'automne pour chacune des régions (*large_area* : Nord-North, Centre-Middle, Sud-South) :

$$M_0 \leftarrow stats :: lm(\log(ce) \sim factor(year) + factor(week), data = cpue_df)$$

et le modèle suivant pour le printemps :

```
M0 ← stats :: lm(log(ce) ~ factor(year) + factor(week), data = cpue_df,  
weight = wt)
```

Le fichier *cpue_models_df_YYYY.csv* contient les estimations des coefficients, des intervalles de confiance 5 % et 95 %, des R carrés et des AIC de ces quatre modèles de régressions. *YYYY* est égale à *yr_global*.

Ensuite, le script calcule la moyenne des moindres carrés (*lsmeans*) pour chaque région et année pour l'automne et pour chaque année pour le printemps et fait la rétro-transformation du logarithme naturel de la moyenne des moindres carrés des CPUE et de leurs intervalles de confiances (5% et 95%) en CPUE kg par filet levé (automne) ou filet (printemps). Le fichier *cpuetran_YYYY.csv* contient ces valeurs.

Les assumptions des modèles, soient : l'homogénéité des variances, l'indépendance des facteurs et la normalité des résidus sont aussi testées. Les graphiques sont produits pour les trois régions pour l'automne et par année pour le printemps et; sont exportés dans le fichier *graphs_reg_assumptions_YYYY.pdf*.

Les fichiers sont entreposés dans le dossier *r_output*.

RÉSULTAT

Cette section décrit les différences de procédures entre les programmes SAS et R et compare les résultats obtenus à la suite de l'exécution de chacune des sections.

LANCEMENT MANUEL DE L'EXÉCUTION AUTOMATIQUE DES SCRIPTS

La structure du flux des opérations des scripts R est un peu différente de celle du flux des opérations des scripts SAS (voir APPENDICE 1). Le lancement du flux des opérations R est manuel, mais l'ensemble des opérations pour produire les valeurs de CPUE est automatique. L'exécution de chacun des scripts SAS est lancée manuellement.

LES OUTILS

fix_net_to_fathoms_fct(x, avg)

Les scripts SAS qui manipulent les données d'effort : *5-Calc number of nets phone dmp_II.sas* et *6-effind_II.sas*, possèdent tous les deux quelques lignes pour effectuer des corrections dans la longueur des filets. Pour uniformiser cette procédure, la fonction *fix_net_to_fathoms_fct* a été créée, mais elle a nécessité quelques ajustements. Les valeurs entre 41 et 49 et entre 57 et 59 n'existent pas dans les fichiers de données entre 1986 et 2021, mais si elles étaient présentes, elles seraient retirées par le programme SAS. La fonction R les assigne à la valeur standard de l'année à laquelle les valeurs appartiennent.

Utilité de la fonction *fix_net_to_fathoms_fct*

Pour les données du DMP, les données brutes de la variable *net_length_fm* datant d'avant 1999 ne sont plus disponibles. Depuis 2003, cette fonction soit principalement utilisée pour assigner la valeur moyenne standard aux données non applicables (NA; tableau 2). Pour le relevé téléphonique, la fonction est utilisée principalement pour les données non applicables et les données en verge (tableau 3).

Tableau 2. Nombre d'observations par classe de taille évaluées par la fonction qui vérifie si la longueur d'un filet est en brasse *fix_net_to_fathoms_fct* pour le programme de monitorages à quai.

year	NA_0:9	10:19	20:40	50:56	60:120	121:500
1999	5015	NA	NA	NA	NA	NA
2000	12179	NA	NA	NA	NA	NA
2001	3070	40	1199	NA	30	NA
2002	8473	2	1674	NA	28	NA
2003	7644	1990	3	39	1	NA
2004	9579	172	18	6	9	10
2005	3034	925	1	22	375	93
2006	7093	2416	10	2	NA	NA
2007	5911	2350	17	4	1	NA
2008	4803	2373	5	3	NA	NA
2009	6652	1407	8	NA	NA	NA
2010	7660	2323	NA	NA	NA	NA
2011	5871	2286	1	1	1	NA
2012	4750	1882	NA	NA	NA	NA
2013	4153	1411	2	NA	NA	NA
2014	3074	1251	2	1	NA	NA
2015	3359	1134	1	NA	NA	NA
2016	3660	1104	7	NA	NA	NA
2017	3704	1188	3	NA	2	NA
2018	3024	713	13	NA	NA	NA
2019	2460	857	4	NA	NA	NA
2020	1842	476	1	NA	NA	NA
2021	1678	508	NA	NA	NA	NA

Tableau 3. Nombre d'observations par classe de taille évaluées par la fonction qui vérifie si la longueur d'un filet est en brasse *fix_net_to_fathoms_fct* pour le relevé téléphonique.

year	NA_0:9	10:19	20:40	50:56	60:120	121:500
1986	76	245	84	22	8	2
1987	33	298	135	15	2	3
1988	8	310	141	5	3	1
1989	6	345	127	3	3	NA
1990	22	431	144	1	1	NA
1991	25	331	111	1	6	NA
1992	26	360	120	1	NA	NA
1993	23	304	107	1	NA	NA
1994	29	265	65	13	3	1
1995	15	259	47	11	2	1
1996	13	334	59	12	NA	NA
1997	55	297	36	2	NA	1
1998	16	269	28	3	NA	NA
1999	58	282	40	6	2	NA
2000	53	273	43	9	4	NA
2001	22	264	41	2	NA	NA
2002	37	210	27	11	2	NA
2003	16	334	18	1	NA	NA
2004	27	273	15	3	2	NA
2005	6	191	28	NA	2	NA
2006	9	131	14	NA	NA	NA
2007	10	133	7	1	NA	1
2008	4	135	20	1	NA	NA
2009	7	137	25	NA	NA	NA
2010	7	146	15	4	NA	NA
2011	15	208	14	3	1	NA
2012	9	171	7	NA	NA	NA
2013	14	171	9	1	NA	NA
2014	15	127	12	1	NA	NA
2015	11	106	5	NA	NA	NA
2016	11	124	19	1	2	NA
2017	11	105	14	2	1	NA

year	NA_0:9	10:19	20:40	50:56	60:120	121:500
2018	6	217	8	NA	NA	NA
2019	6	240	5	NA	NA	NA
2020	9	186	13	NA	NA	NA
2021	1	65	3	NA	NA	NA

Aires spatiales du hareng du sGSL : *A2-herring_areas.R*

Chacun des scripts SAS qui manipulent les données possède quelques lignes pour effectuer des conversions entre les agrégations et aires spatiales. Il existe trois différences principales entre les définitions des agrégations utilisées par SAS et R. Le script *A2-herring_areas.R* qui produit le fichier *areas.csv* effectue :

- le retrait des districts statistiques (Figure 24 de l'Appendice 2) qui ne font pas partie de l'évaluation des sous-populations de la région du Golfe, soit :
 - les districts 14 et 30 sur la côte sud de la Nouvelle-Écosse;
 - les districts 19, 23, 24 correspondant respectivement à Sept-Îles, La Tabatière et Blanc-Sablon au Québec, et;
 - les districts 72 et 79 correspondant aux eaux intérieures du Nouveau-Brunswick.
- Le changement (SAS vs R) du district 76 au Nouveau-Brunswick de l'agrégation :
 - *unknown_area* 2 à 1 et;
 - *herring_area* 16E à 16C;
 de façon à correspondre mieux à la séparation effectuée par les agrégations *telsurvey_area* 4 et *large_area* Centre (*Middle*; voir Tableau 1).
- L'ajout (SAS vs R) de 4Vn par le changement des districts 1, 4, 6 et 7 de la Nouvelle-Écosse de l'agrégation :
 - *herring_area* 16G (district 1) ou *herring_area* 16F (districts 4, 6 et 7) à *herring_area* 17 (districts 1 et 4) ou *herring_area* 18 (districts 6 et 7);
 - *unknown_area* 8 à 9;
 - *telsurvey_area* 6 à 9, et;
 - *large_area* Sud (South) à 4Vn.

SOURCES DE DONNÉES

Chaque année, les programmes SAS uniformisaient et/ou agrégeaient les données brutes et archivaient les données transformées pour obtenir les valeurs des variables liées à la capture et à l'effort nécessaire au calcul de la CPUE. Ces fichiers étaient utilisés pour effectuer les calculs des années subséquentes. Antérieurement à 2018, il n'existait pas d'archivage systématique des scripts décrivant de quelle façon ces données avaient été modifiées, excluant les données du relevé téléphonique.

DMP : Programme de monitoring à quai

De façon générale, le traitement des données du DMP à l'aide des scripts SAS n'était pas exactement le même entre les années, entre la région du Québec et du Golfe du MPO et entre les saisons de pêche (voir APPENDICE 1 et le dossier *sas_script* du projet R). Ces différences annuelles dans le traitement et l'archivage des données apportaient des complications dans le traitement. Le traitement des données à l'aide des scripts R est uniforme entre les années, les régions et les saisons de pêches, à l'exception des années où les données brutes n'existent plus. Ceci est possible, car c'est une boucle qui prend les fichiers annuels de données brutes et qui les fait passer par une série d'opérations pour effectuer les corrections, et ensuite; ces fichiers sont fusionnés, puis modifiés par un seul script.

Pour les années entre 2007 et 2021, la procédure SAS utilise le script *reform_dmp.sas* qui ne garde pas la variable *trip_id* et rend disponible la variable *herring_area* et *port_landed*. Ce script permet d'obtenir les données de nombre de filets standards, de capture et de nombre de voyages pour les deux régions pour les deux saisons. Le code *merge_dmp_net_lands_data.sas* garde la variable *trip_id* et rend disponible les variables *telsurvey_area* et *port_landed*. Ce script permet d'obtenir les données de nombre de filets standards et de nombre de voyages pour la région du Golfe pour les deux saisons. Il y a un dédoublement de l'information, mais le fichier produit par *reform_dmp.sas* est utilisé pour obtenir le nombre de filets standards seulement pour la région du Québec pour les deux saisons et pour obtenir la capture pour le printemps pour les deux régions; tandis que le script *merge_dmp_net_land_data_II.sas* est utilisé pour obtenir le nombre de filets standards pour la région du Golfe pour les deux saisons. De plus à la suite du calcul du nombre de filets standard à l'aide du script SAS *calc number of nets phone dmp_II.sas*, les données produites pour les années 2005 à 2013 sont arrondies et archivées seulement si les nombres d'observations obtenus sont plus élevés que ceux obtenus par le relevé téléphonique annuel par *telsurvey_area* et *season*. Les données entre 1999 et 2004 ne sont pas disponibles.

Le script *B1-dmp_sas_vs_r.R* permet de produire les

Figure 5. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche printanière, obtenus par le DMP entre 1999 à 2021. (printemps) et

Figure 6. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche automnale, obtenus par le DMP entre 1999 à 2021. R en bleu. SAS en orange. (automne) qui comparent le nombre de filets standards utilisés pour tous les bateaux par *telsurvey_area par saison* obtenus par :

- le programme SAS *calc number of nets phone dmp_II.sas* qui a permis de recréer les données utilisées de 2007 à 2014 et les fichiers *YYYYClean_Gulf_QC_Catch_Effort.RData* pour les données de 2015 à 2021 (Cosandey-Godin 2017, procédure R non publiée) avec;
- le fichier *dmp_std*.

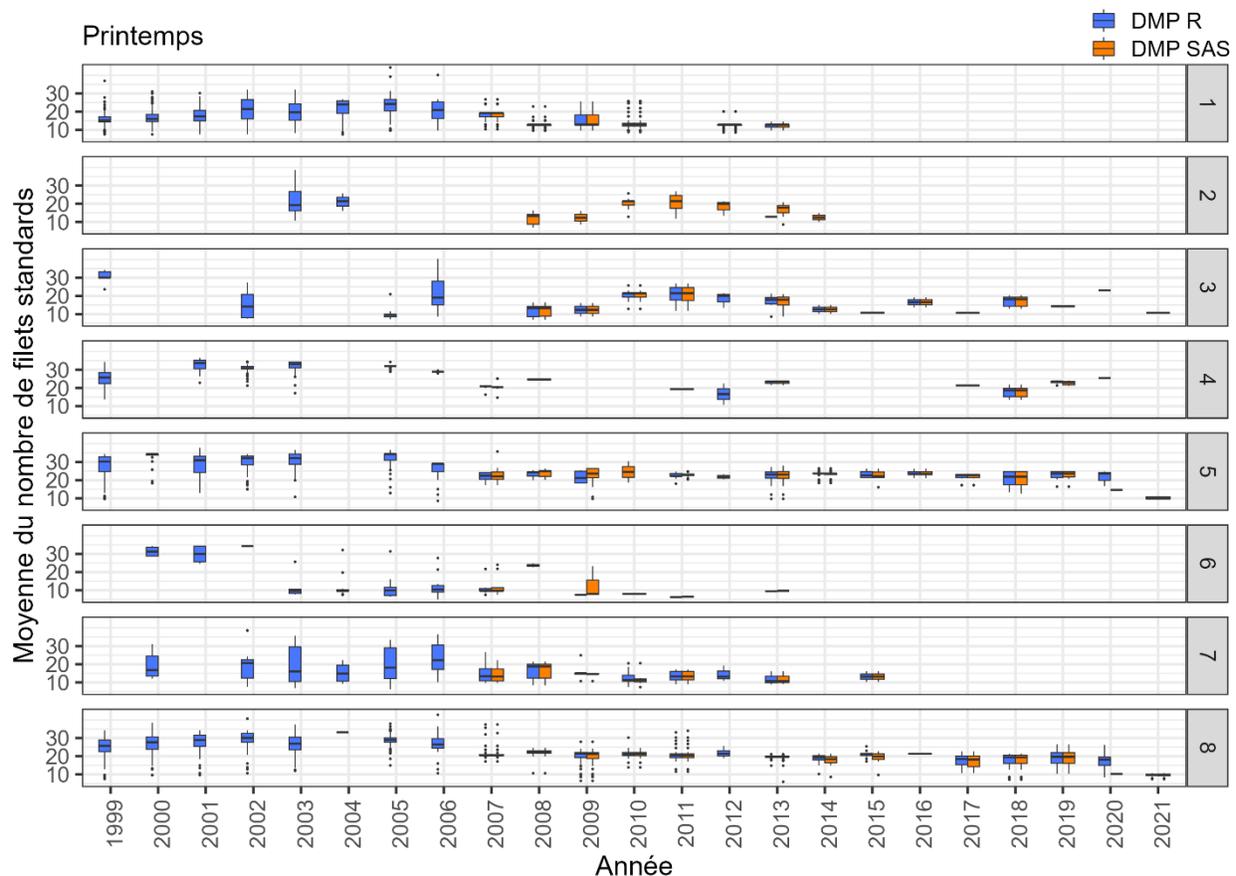


Figure 5. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche printanière, obtenus par le DMP entre 1999 à 2021. R en bleu. SAS en orange.

Une observation représente une moyenne de toutes les valeurs obtenues par le DMP du nombre de filets standards (*nets_std*) pour un bateau (*cfv*) dans une année (*year*), pour une saison (*season*), par *telsurvey_area*. Principalement, on observe qu'il y a pas de données fournies par la procédure SAS avant 2007. D'autres différences sont reliées à la correction de certaines assignations aux agrégations du relevé téléphonique (*telsurvey_area*) à l'aide des port de débarquement ou à l'ajout de ports de débarquement manquants. Par exemple, on peut observer une présence de données pour la procédure SAS et une absence de données dans R pour la pêche printanière à la *telsurvey_area* 2 entre 2008 et 2013. Les bateaux (*cfv*) correspondant à cette *telsurvey_area* 2 dans le fichier produit par SAS pour ces années sont assignées aux *telsurvey_area* 3 et 5 dans le fichier produit par la procédure R. Il y a une possibilité que ces évènements de pêche de l'aire 2 aient eu lieu au Québec et aient été débarqués au Nouveau-Brunswick.

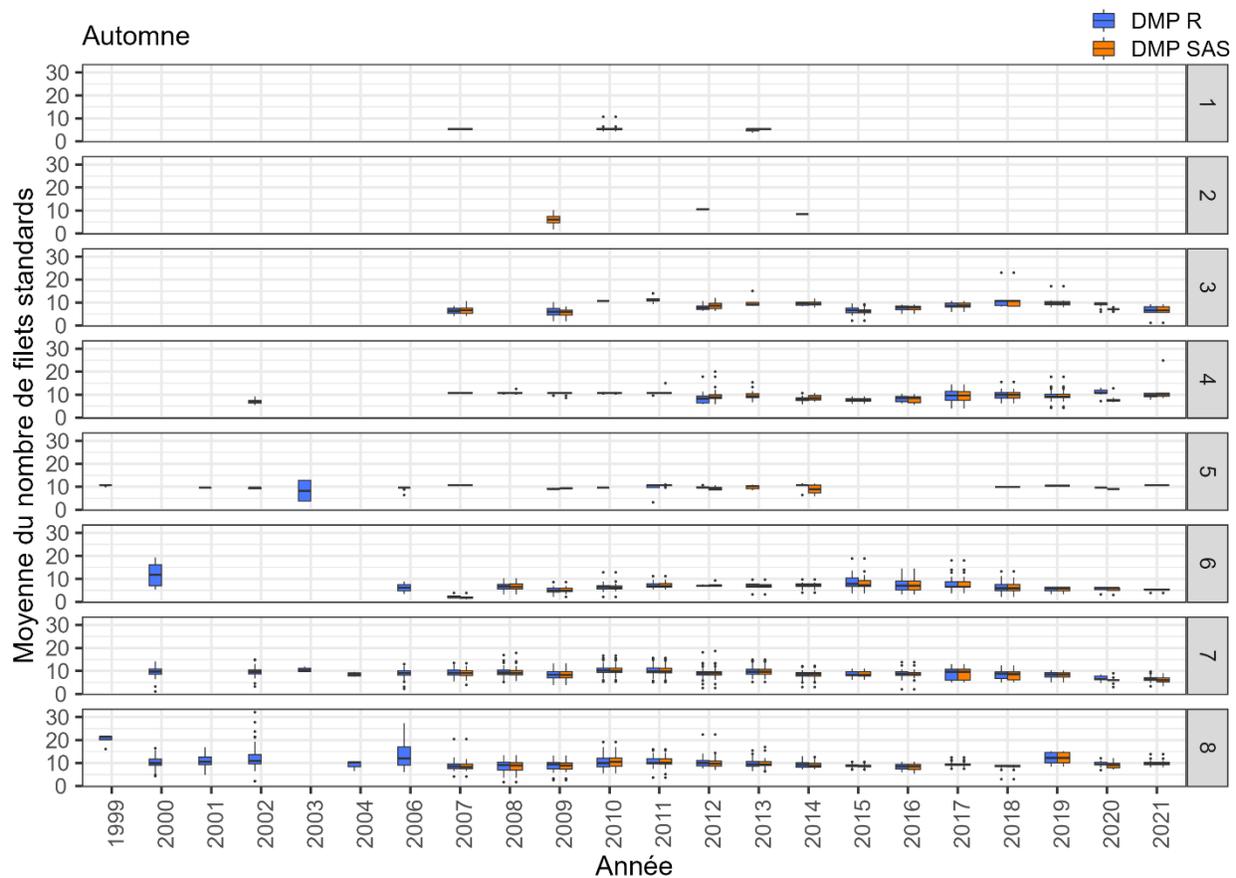


Figure 6. Nombre moyen de filets standards (boîte à moustache) par bateau, par année et par aire du relevé téléphonique pour la pêche automnale, obtenus par le DMP entre 1999 à 2021. R en bleu. SAS en orange.

La Figure 7 compare les résultats obtenus par année pour la capture *tot_catch_kg* et le nombre de voyages *trips* pour obtenir cette capture, par le programme SAS *cpue_spring_ll.sas* et le programme R *1d-dmp.R* qui produit le fichier *dmp_catch*. Une figure est présentée pour le printemps seulement puisque les données du DMP ne sont pas utilisées pour obtenir les valeurs de captures et de nombres de voyages pour l'automne. De plus, les valeurs de captures colligées par le DMP proviennent seulement de quelques districts statistiques, les valeurs pour les autres districts proviennent des ZIFF (voir section ZIFF). On remarque que les différences observées dans les nombres de voyages et les captures se retrouvent après 1998 et que ces différences sont plus fréquentes après 2001.

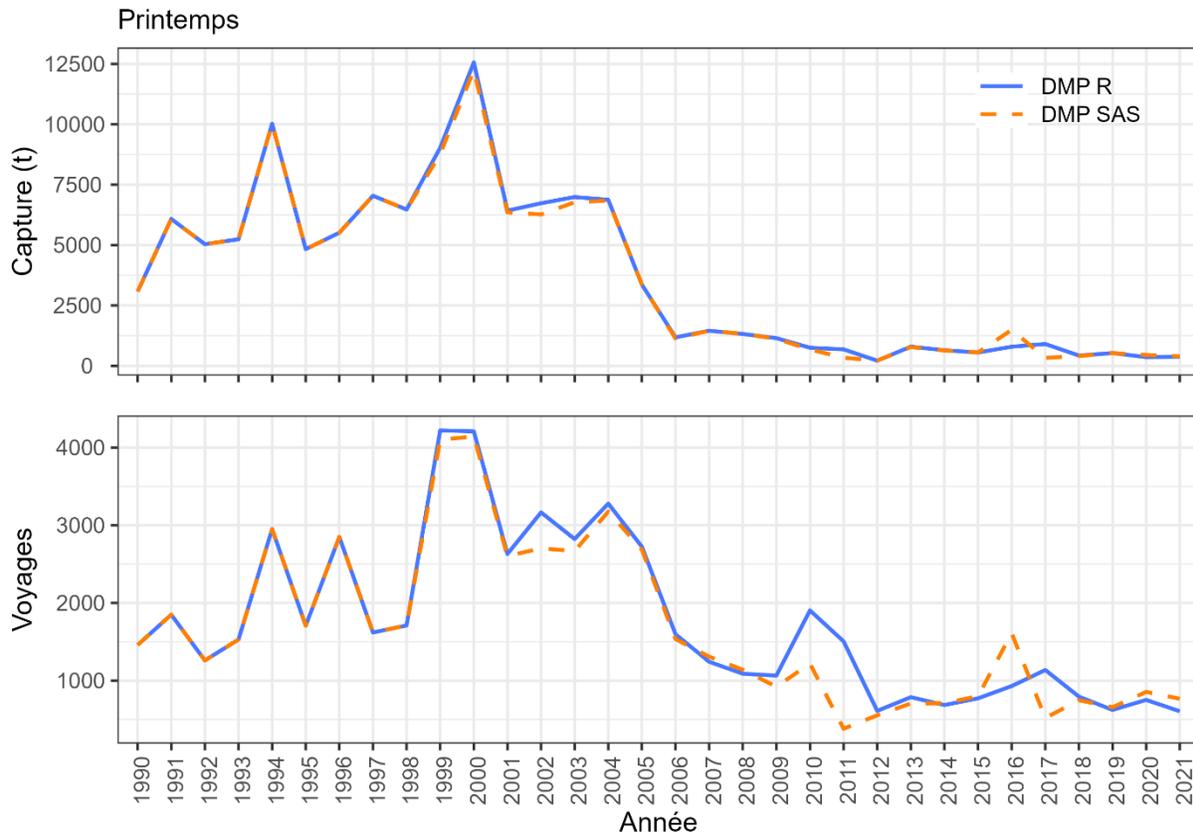


Figure 7. Capture en tonne et nombre de voyages pour la pêche printanière, enregistrés par le DMP entre 1990 à 2021. R en bleu. SAS en orange.

Telsurvey : Relevé téléphonique

Pour le relevé téléphonique, il n'y a pas de différence particulière entre la procédure SAS et R (voir APPENDICE 1) pour la manipulation des données. Cependant, certaines erreurs dans la procédure SAS ont été corrigées dans celle de R lors du calcul d'un nombre moyen de filets standard *avgnets*.

Les scripts R et SAS vérifient si la pêche a lieu durant une saison où il y a un pic d'abondance ou non (*peak* égale à Y ou N). Certaines lignes du fichier de données n'ont pas d'information pour cette variable. La procédure SAS vérifie si *peak* égale à Y et s'il ne l'est pas, il est égal à N, ce qui n'est pas toujours le cas. Dans la procédure R, lorsque *peak* égale à NA et que la valeur du *peakdays* (le nombre de jours de pêche où on observe une forte abondance) est présente, *peak* est assignée à Y.

Lorsque *peak* égale Y les procédures SAS et R déterminent combien de jours il y avait sans pic d'abondance soit : *npdays* est plus grand que 0. Normalement, *npdays* égale à *totdays* (le nombre total de jours de pêche) moins *peakdays*. Cependant, il arrive que *totdays* ne soit pas comptabilisé et que *peakdays* le soit ou que *peakdays* soient plus élevés que *totdays*. Dans ce cas, le programme SAS obtient un résultat négatif et

l'assigne à 0. Le programme R assigne le plus grand nombre à *totdays* et donc n'obtiens jamais de valeur négative.

Le script *B2-telsurvey_sas_vs_r.R* permet de produire des graphiques comparatifs des résultats obtenus par le fichier SAS *next1* produit par le programme SAS *effind_ll.sas* et R *tel_cfv* produit par le programme R *4-effort.R*. Les Figure 8 et Figure 9 comparent les nombres de filets standard obtenus par année (*year*), par bateau (*cfv*) ou responsable de bateau (*resp*) et pour la saison (*season*) de printemps et d'automne respectivement.

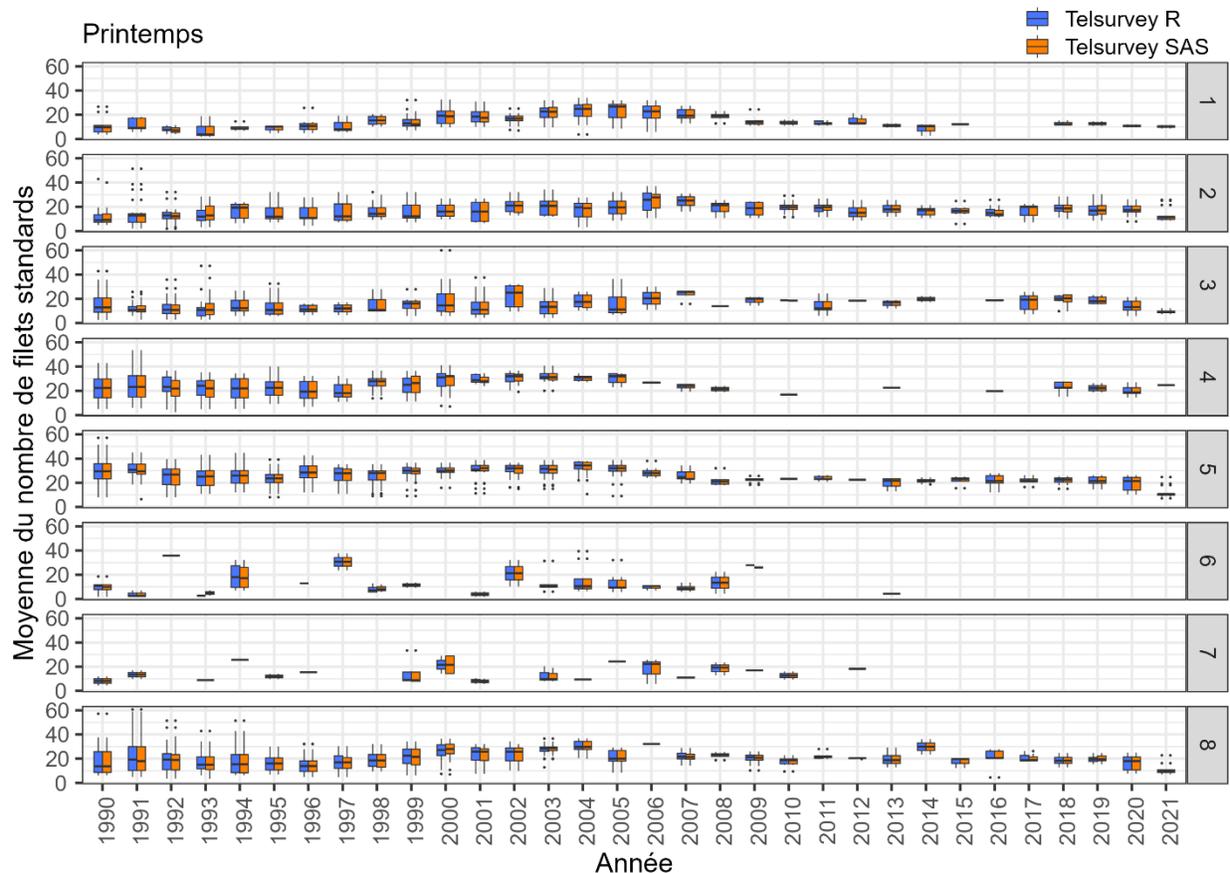


Figure 8. Nombre moyen de filets standards *avg_nets* (boîte à moustaches) pour la pêche printanière par aire du relevé téléphonique *telsurvey_area* entre 1990 et 2021. R en bleu. SAS en orange.

Les données agrégées pour fin de comparaison ne proviennent pas du fichier *telsurvey_std.R*, car l'agrégation est une partie du calcul de l'effort et est effectuée dans le programme *effind_ll.sas*. Il y a agrégation, car le programme *4-effort.R* effectue une moyenne des nombres moyens de filets standard pour obtenir une seule information par bateau ou responsable par année et aire du relevé téléphonique. Le programme SAS ne fait pas cette opération, il est donc possible d'observer plus qu'une observation par bateau, responsable ou responsable et bateau et le résultat obtenu pour nombre moyen de filets standard peut être différent pour certaines aires et années entre les deux procédures.

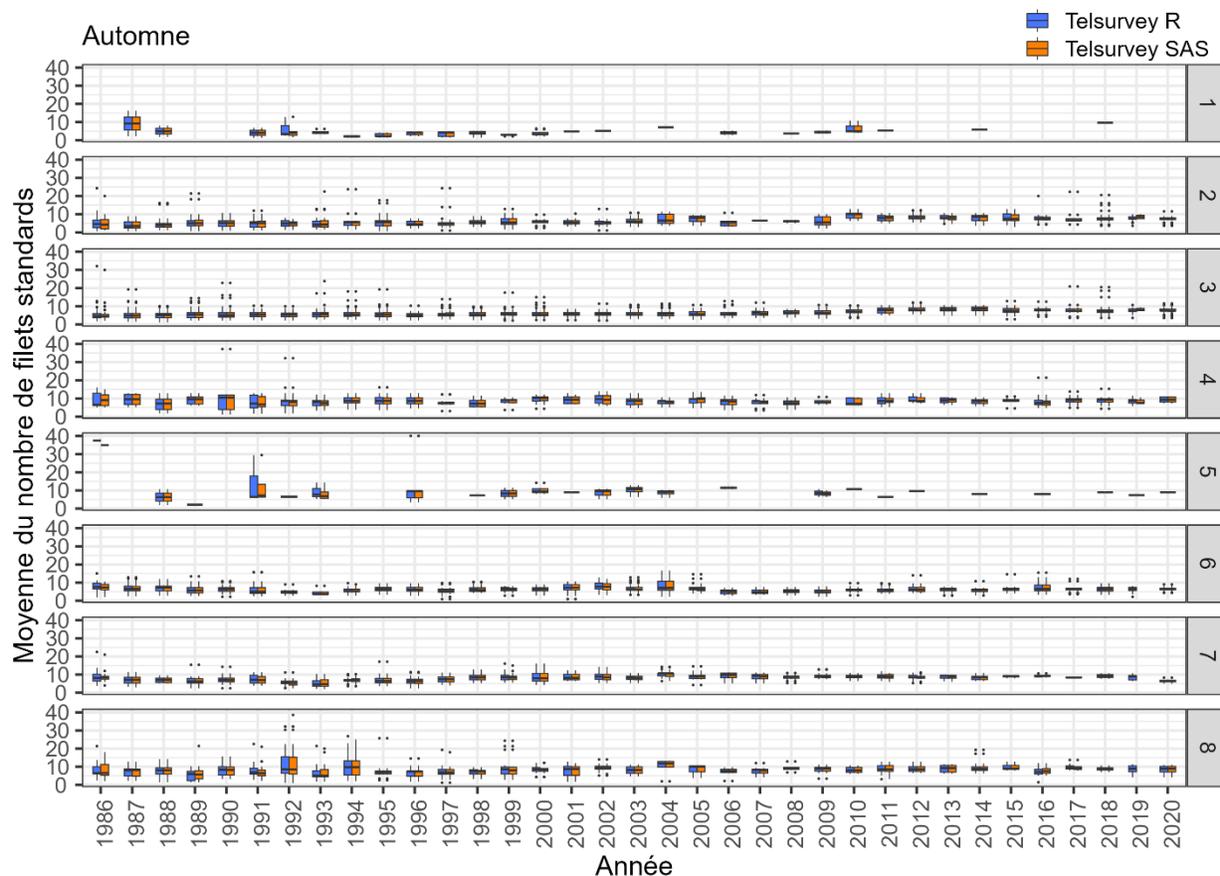


Figure 9. Nombre moyen de filets standards *avg_nets* (boîte à moustaches) pour la pêche automnale par aire du relevé téléphonique *telsurvey_area* entre 1990 et 2021. R en bleu. SAS en orange.

Malgré cette différence, celles dans la définition des aires (*A2-herring_areas.R*) et les différences dans la méthodologie du calcul de *avgnets* énoncées plus haut, il ne semble pas avoir de différence majeure entre les deux procédures.

ZIFF : fichier informatisé inter-zone

Pour les ZIFF, les données brutes sont entreposées dans des bases de données *Hd3:/landings/zif/raw* (1986 à *yr_global*) et *Hd3:/landings/pslip* (1986 à 2002, région du Golfe seulement). Le script SAS *hergrade_II.sas* importe les données brutes et les transforme. Les fichiers de données produits des années antérieures : *hercpue_YYYY.dat*, sont réutilisés pour l'évaluation de l'année en cours. Il n'y a pas ou peu de traces de ce qui a été fait sur les données brutes des années antérieures à 2018.

Le script *B3-zif_sas_vs_r.R* effectue la comparaison entre les procédures SAS et R. Quelques problèmes ont été décelés lors de la manipulation des données pour traduire le code SAS en R.

- Pour la sélection des numéros de bateau (*cfv*), le script SAS *8-hergrade_II.sas* utilise et sélectionne les données sous le format de caractères ("000001" à

“189999”) et donc sélectionne seulement les numéros qui possèdent 6 caractères et les omet dans le cas contraire. Il y a des modifications aux codes pour corriger cette sélection annuellement, mais les corrections n’ont pas été effectuées de façon systématique. La procédure R utilise le format numérique ce qui permet de tous les intégrer.

- Les données archivées contiennent parfois des doublons qui ne sont pas retirés par la procédure SAS.
- Les données de 1986 à 2002 semblent en partie formées des données de bon d’achat au débarquement : Pslips. Si on utilise seulement les données des ZIFF pour les années entre 1986 et 2002, il n’est pas possible d’obtenir une bonne reproductibilité entre les résultats obtenus avec les données produites par SAS et archivées annuellement et les résultats obtenus par la procédure R.

Les Figure 10 et Figure 11 montrent la capture en kg et le nombre de voyages issus des ZIFF et/ou des Pslips utilisées pour les calculs de la capture en kg pour le printemps et l’automne respectivement. Ces figures permettent aussi de retrouver à partir de quelle source de données probables les données des ZIFF ont été construites entre 1986 et 2002.

Il n’est pas possible de retrouver la procédure qui fusionne les Pslips aux ZIFF utilisés par le programme SAS puisque les scripts utilisés annuellement entre 1986 et 2002 n’ont pas été archivés systématiquement. Cependant, la procédure utilisée par les programmes R est la même pour la pêche printanière et automnale. On remarque qu’il n’est pas possible d’obtenir une bonne reproductibilité des résultats annuellement pour la pêche printanière, mais que la procédure est probablement adéquate étant donné la bonne reproductibilité obtenue pour la pêche automnale.

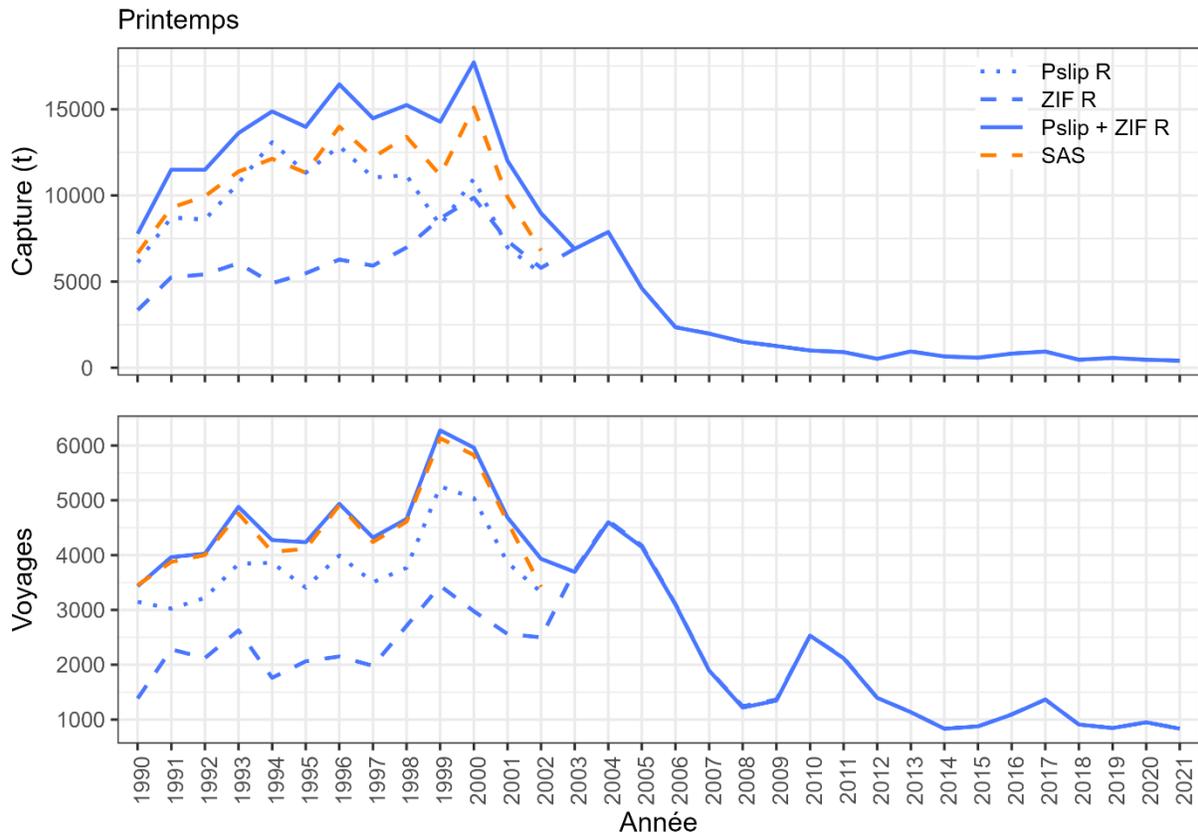


Figure 10. Capture en tonne et le nombre de voyages pour la pêche printanière enregistrés par les ZIFF entre 1990 et 2021 (hachuré), Pslips (pointillé) et la combinaison des ZIFF et des Pslips (trait complet). R en bleu. SAS en orange.

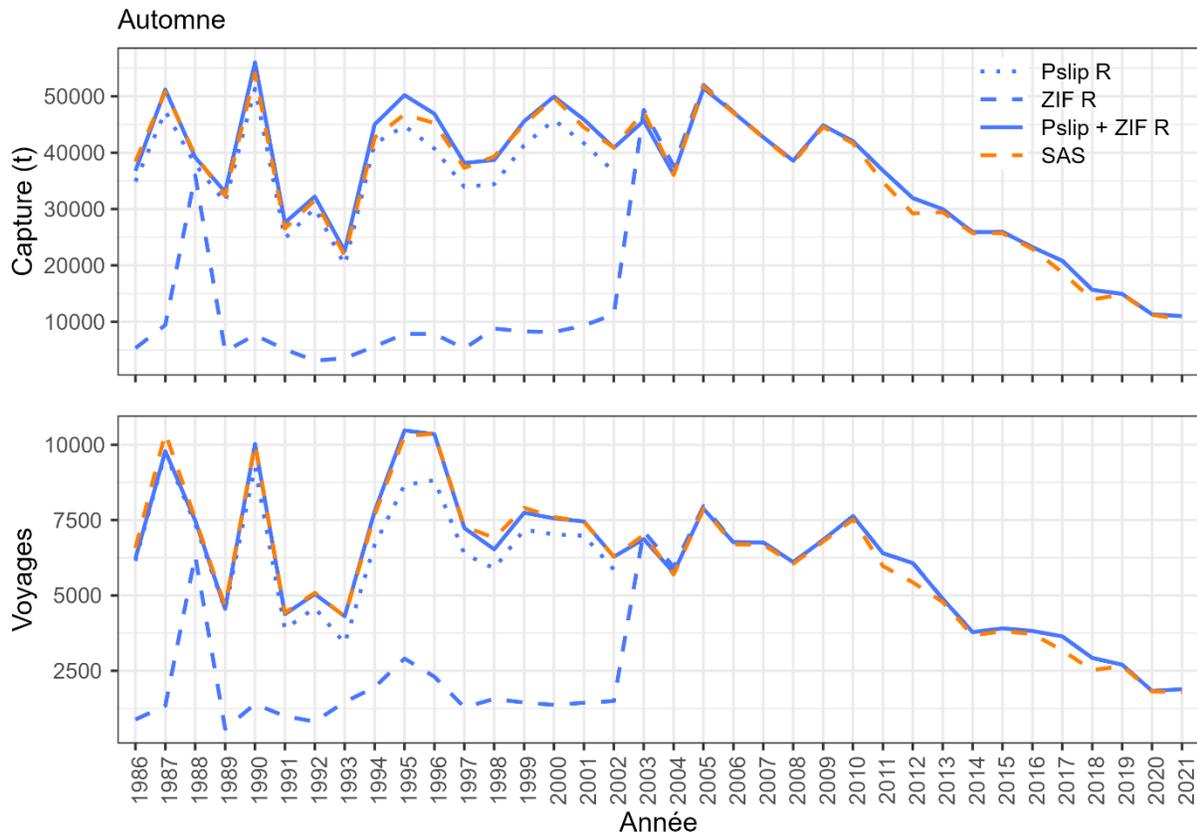


Figure 11. Capture en tonne et le nombre de voyages pour la pêche automnale enregistrés par les ZIFF entre 1990 et 2021 (hachuré), Pslips (pointillé) et la combinaison des ZIFF et des Pslips (trait complet). R en bleu. SAS en orange.

CALCULS

Effort

Le nombre moyen de filets standardisés est obtenu par l'agrégation des données par année, saison, aire de relevé téléphonique, numéro de bateau ou responsable pour les données du relevé téléphonique et du programme de monitoring à quai.

Comme mentionné précédemment, la procédure SAS pour le relevé téléphonique n'effectue pas de moyenne par bateau et/ou responsable, bien que plusieurs valeurs peuvent être observées contrairement à la procédure R.

Le choix est effectué entre les deux relevés de façon différente entre la procédure SAS et R. De façon générale pour les deux procédures, il s'agit d'une procédure séquentielle où pour chacun des relevés la première valeur de nombre moyen de filets standardisés par année est écrasée par la suivante, soit : le nombre moyen de filets standardisés par année et agrégation du relevé téléphonique, si elle existe. La valeur d'un relevé est choisie si elle obtient un nombre d'observations supérieur à l'autre relevé. Cependant, la séquence des opérations pour la procédure SAS effectuée par les

programmes *10-cpue_spring_ll.sas* et *9-cpue_fall_hauls_ll.sas* est un peu différente de la procédure R et de ce qui est attendu à la suite de la lecture des documents de recherche, la voici :

1. Un fichier vide comprenant les variables : année (1986 à *yr_global*), saison (printemps, automne), aires du relevé téléphonique (1 à 8), est créé;
2. Les nombres moyens de filets standardisés par saison, par année, pour le relevé téléphonique seulement (1986 à *yr_global*) sont ajoutés;
3. Les nombres moyens de filets standardisés par aires du relevé téléphonique, par saison, par année, pour le relevé téléphonique seulement (1986 à *yr_global*) sont ajoutés;
4. Les nombres moyens de filets standardisés par aires du relevé téléphonique, par saison, par année, calculés annuellement pour le DMP par le programme *5-calc number of nets phone dmp_ll.sas*, s'ils obtiennent un nombre d'observations supérieur à ceux du relevé téléphonique et; archivés à l'intérieur des programmes pour le printemps et l'automne correspondant, sont ajoutés (voir APPENDICE 1).
5. Les données du nombre de levées multipliées par le nombre de filets *eff2_mean* par aire du relevé téléphonique, par saison, par année (1986 à *yr_global*), en provenance des relevés téléphoniques, sont ajoutées s'il existe une valeur par aire du relevé téléphonique et année.

Ainsi, Les valeurs du DMP y sont archivées depuis 2005 pour le printemps et depuis 2006 pour l'automne bien que des valeurs soient disponibles depuis 1999 pour certaines aires du relevé téléphonique. De plus, il n'y a pas de choix de la valeur du nombre de filets entre les relevés pour l'agrégation par saison et année. Enfin, la procédure R est automatisée avec l'ensemble des autres scripts alors que les données doivent être coupées et collées manuellement dans les scripts *10-cpue_spring_ll.sas* ou *9-cpue_fall_hauls_ll.sas* pour le printemps et l'automne respectivement pour la procédure SAS.

Les Figure 12 et Figure 13 montrent la différence entre les procédures SAS et R pour le nombre moyen de filets standard *nets_std_mean* pour la capture de printemps et d'automne. La Figure 14 montre la différence entre la procédure SAS et R pour le nombre de filets relevés *eff2_mean* choisi pour la capture d'automne.

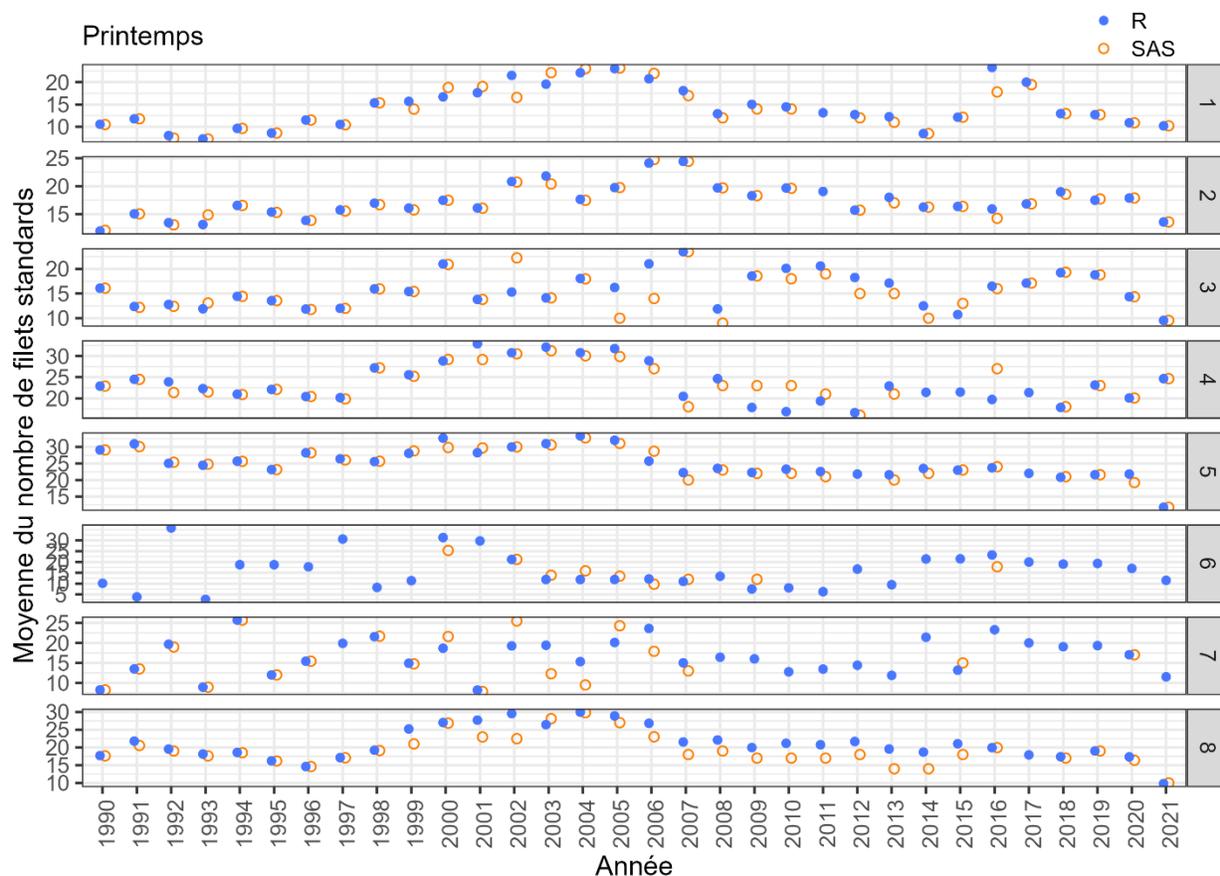


Figure 12. Nombre moyen de filets standards utilisés pour la pêche printanière par aires du relevé téléphonique (1 à 8) entre 1990 et 2021. R en bleu. SAS en orange.

On remarque une similarité entre les résultats issus des 2 procédures, mais aussi que l'automatisation et les modifications à la procédure permettent d'obtenir une valeur pour toutes les données manquantes. Principalement, la procédure R permet d'obtenir des valeurs pour la pêche printanière pour les agrégations du relevé téléphonique 6 et 7 où une grande partie des observations sont manquante et, pour l'année 2017 où des observations sont manquantes pour les agrégations du relevé téléphonique 4 à 8. Pour la pêche d'automne, la procédure R complète les trouées, principalement pour l'aire du relevé téléphonique 1 et, pour l'année 2019 où des observations sont manquantes pour les agrégations du relevé téléphonique 1 et 5 à 8.

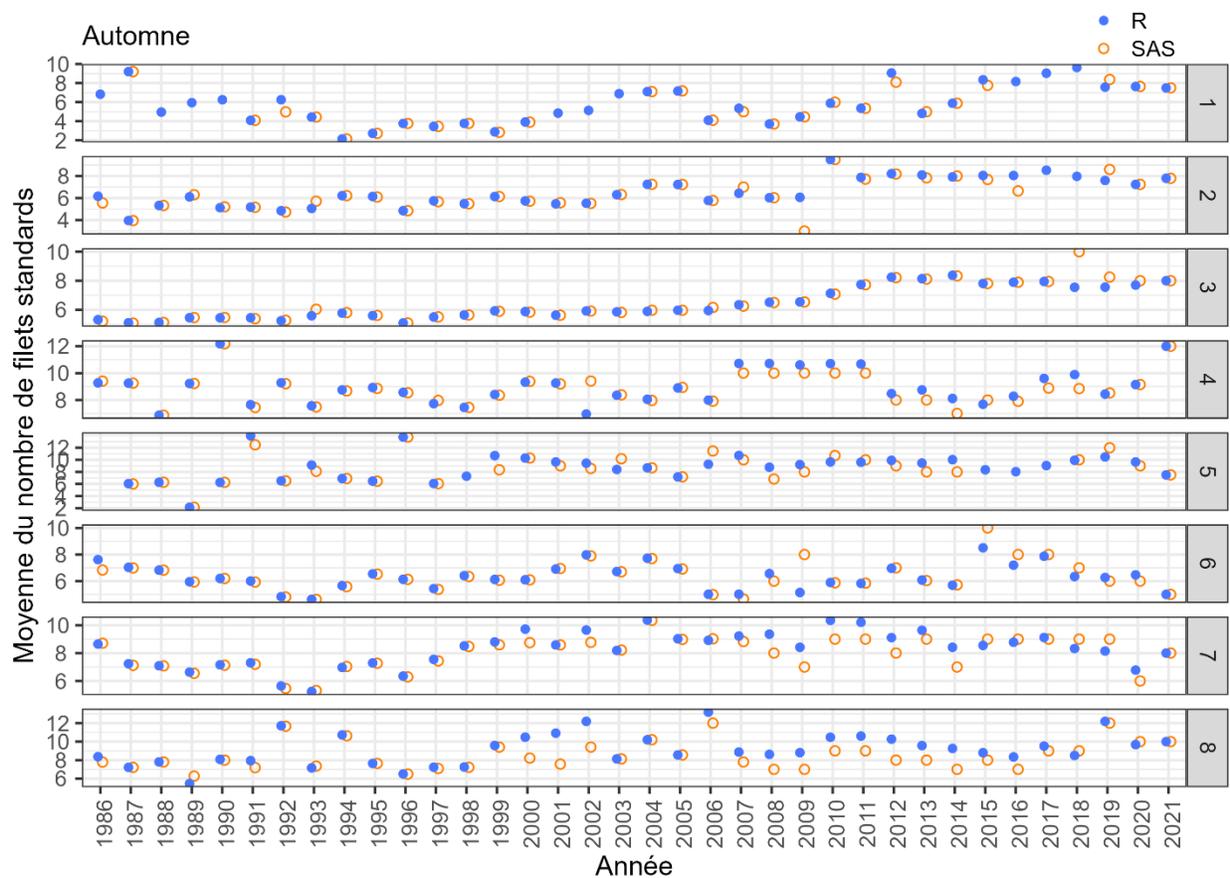


Figure 13. Nombre moyen de filets standards utilisés pour la pêche automnale par aires du relevé téléphonique (1 à 8) entre 1986 et 2021. R en bleu. SAS en orange.

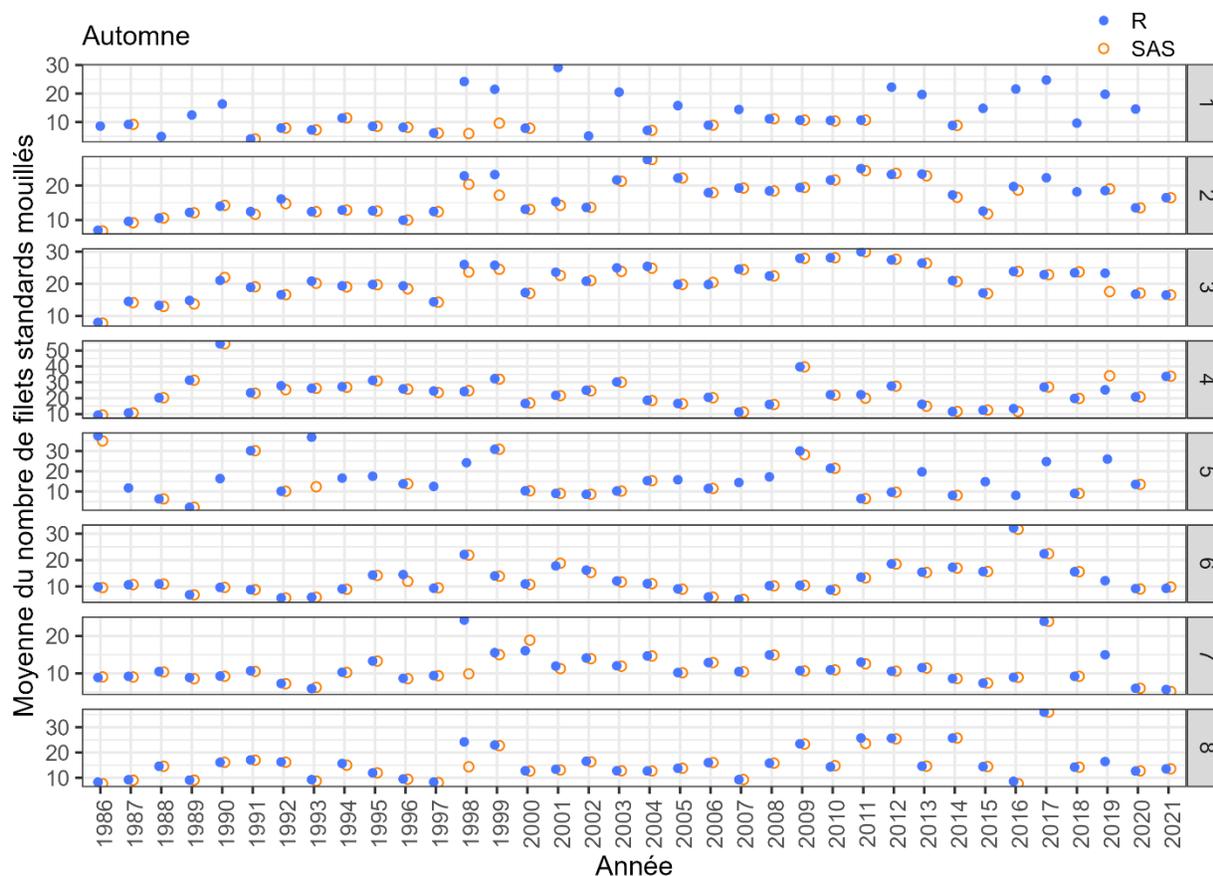


Figure 14. Nombre moyen de filets standards mouillés ou relevés pour la pêche automnale par aires du relevé téléphonique (1 à 8) entre 1986 et 2021. R en bleu. SAS en orange.

Capture

Le calcul de la capture diffère entre les programmes SAS et R pour le printemps seulement. La procédure SAS utilise les ZIFF pour compléter les données jusqu'en 2002 (*hercpue_1986_2002.dat* et *hercpueque_1985_2002.dat*), mais elle retire les données de ZIFF appartenant aux *unknown_area* : 1, 2 et 4 pour 1999; 1, 2 et 3 pour 2000 et 2002, et; 4 pour 2002 seulement. Pour les années suivantes (2003 à 2021), la procédure utilise seulement les DMP. Le script R prend les données des ZIFF provenant des districts qui ne sont pas présents dans les données du DMP pour toute la série temporelle (1990 à *yr_global*) tels que décrit dans les rapports de recherches reliés aux évaluations de stock de hareng du sGSL (Rolland et al. 2022).

La comparaison des résultats obtenus par les programmes SAS *10-cpue_spring_ll.sas* et *9-cpue_fall_hauls_ll.sas* et R *5-catch.R* est effectuée dans le programme *B5-catch_sas_vs_r.R*. Les Figure 11 et Figure 15 montrent les résultats de captures calculées par les programmes SAS et R pour l'automne et le printemps respectivement. On voit une différence entre les résultats obtenus entre les résultats R et SAS, surtout pour les années antérieures à 2003.

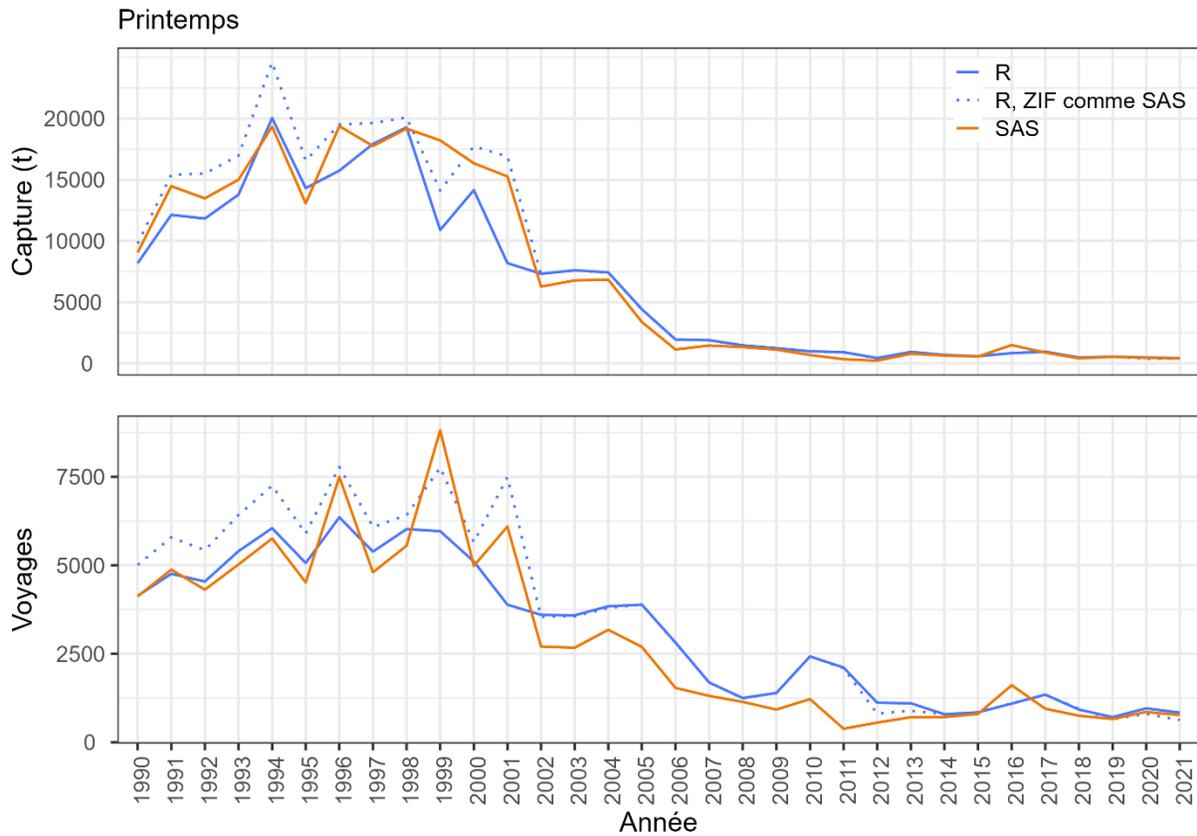


Figure 15. Capture en tonne et du nombre de voyages pour la pêche printanière enregistrés par le DMP et complétés par les ZIFF entre 1990 et 2021. R en bleu. R avec procédure SAS de sélection des ZIFF en bleu pointillé. SAS en orange.

CPUE

La procédure SAS et R sont similaire pour le calcul de la CPUE.

Cependant, à cette étape pour la pêche printanière les données assignées aux aires 1 à 8 du relevé téléphonique sont sélectionnées et pour la pêche automnale les données assignées aux régions Sud, Centre et Nord le sont. On remarque un nombre d'observations qui est utilisées par année, mois, jour et district pour le calcul est plus élevé pour la procédure R (printemps : 19 879, automne : 15 713) que pour celle de SAS (printemps : 9 598, automne : 15 117, Figure 16 et Figure 17).

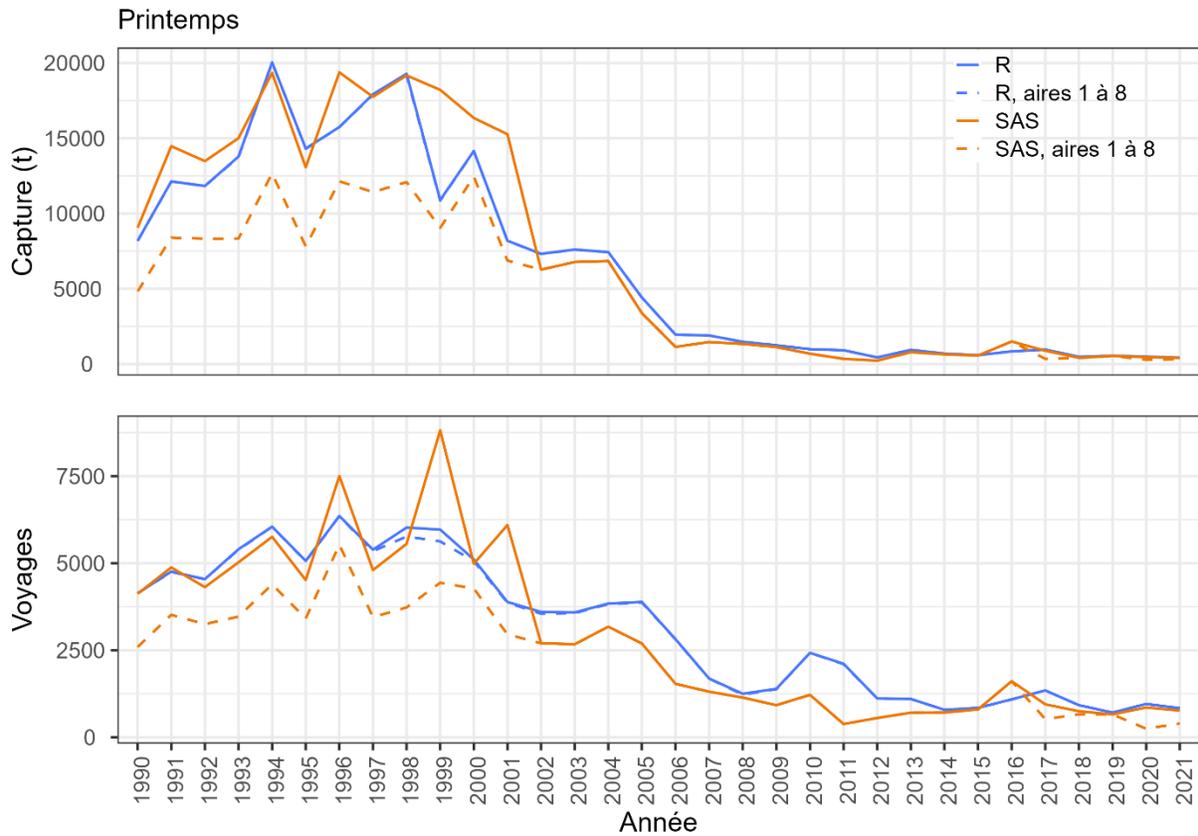


Figure 16. Capture en tonne et du nombre de voyages pour la pêche printanière enregistrés par le DMP et complétés par les ZIFF entre 1990 et 2021. R en bleu. SAS en orange. Procédures avec (trait hachuré) et sans (trait plein) le retrait des *unknown_area* non applicable et égale à 9.

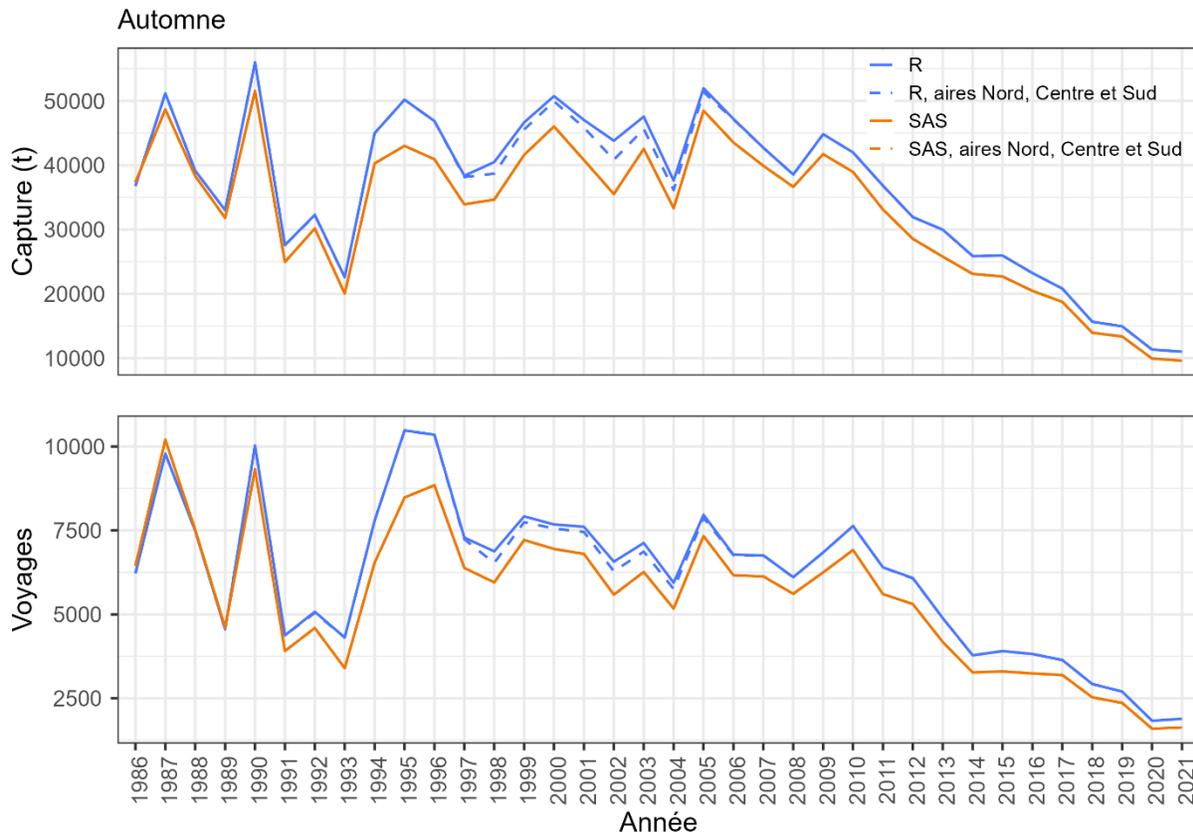


Figure 17. Capture en tonne et du nombre de voyages pour la pêche automnale enregistrés par les ZIFF entre 1986 et 2021. R en bleu. SAS en orange. Procédures avec (trait hachuré) et sans (trait plein) le retrait des *large_area* non applicable et égale à $4\sqrt{n}$.

La comparaison des résultats obtenus par les programmes SAS *10-cpue_spring_ll.sas* et *9-cpue_fall_hauls_ll.sas* avec; R *6a-cpue_data.R* et *6a-cpue_model.R* est effectuée dans le programme *B6-cpue_sas_vs_r.R*. Les Figure 18 et Figure 20 montrent la variabilité des données de CPUE qui entrent dans le modèle et les Figure 19 et Figure 21 montrent les estimations de la CPUE des modèles, pour le printemps et l'automne respectivement. Le nombre d'observations agrégé qui entre dans le modèle du printemps et de l'automne est aussi plus élevé pour la procédure R (printemps : 3 077, automne : 1 042) que celle de SAS (printemps : 2 601, automne : 854).

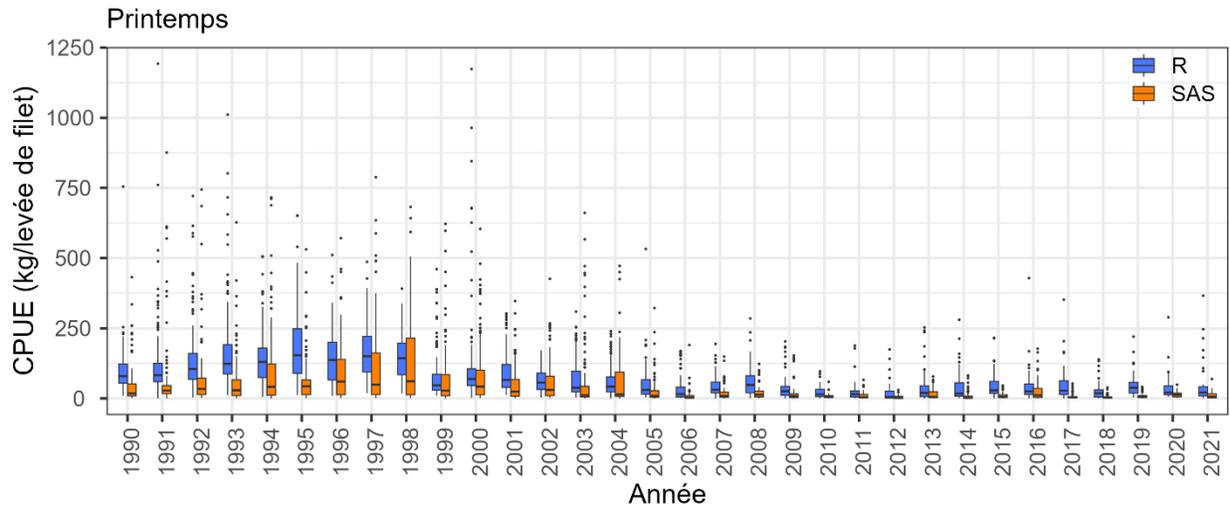


Figure 18. CPUE calculées (boîte à moustaches) par aire de pêche au hareng divisé par province (*unknown_area*), par semaine et par jour julien, pour la pêche printanière entre 1990 et 2021. R en bleu. SAS en orange.

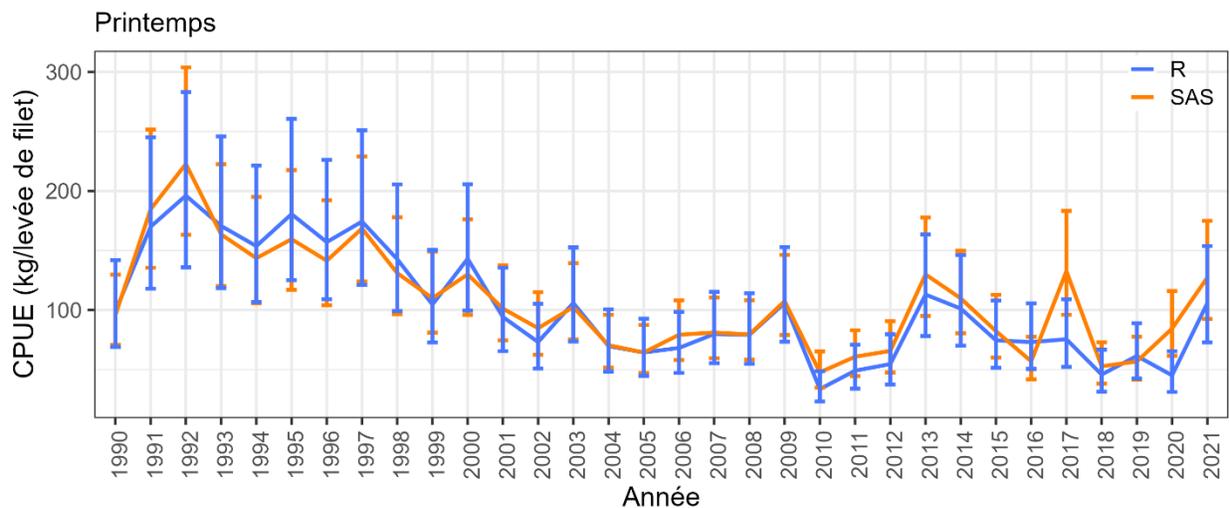


Figure 19. Estimations de la CPUE (kg par levée de filet, moyenne et intervalles de confiance 5 et 95 %) pour la pêche printanière entre 1990 et 2021. R en bleu. SAS en orange.

Pour la pêche printanière, la Figure 18 montre que les valeurs des données entrantes dans le modèle sont plus élevées lorsqu'elles proviennent de la procédure R que de la procédure SAS. On remarque aussi à la Figure 19 une grande différence entre les deux procédures pour l'année 2017. Différence probablement due à la complétion des données de l'effort particulièrement pour l'année 2017, voir Figure 12, à la très grande différence entre le nombre d'observations qui entre dans le modèle pour R comparativement à SAS.

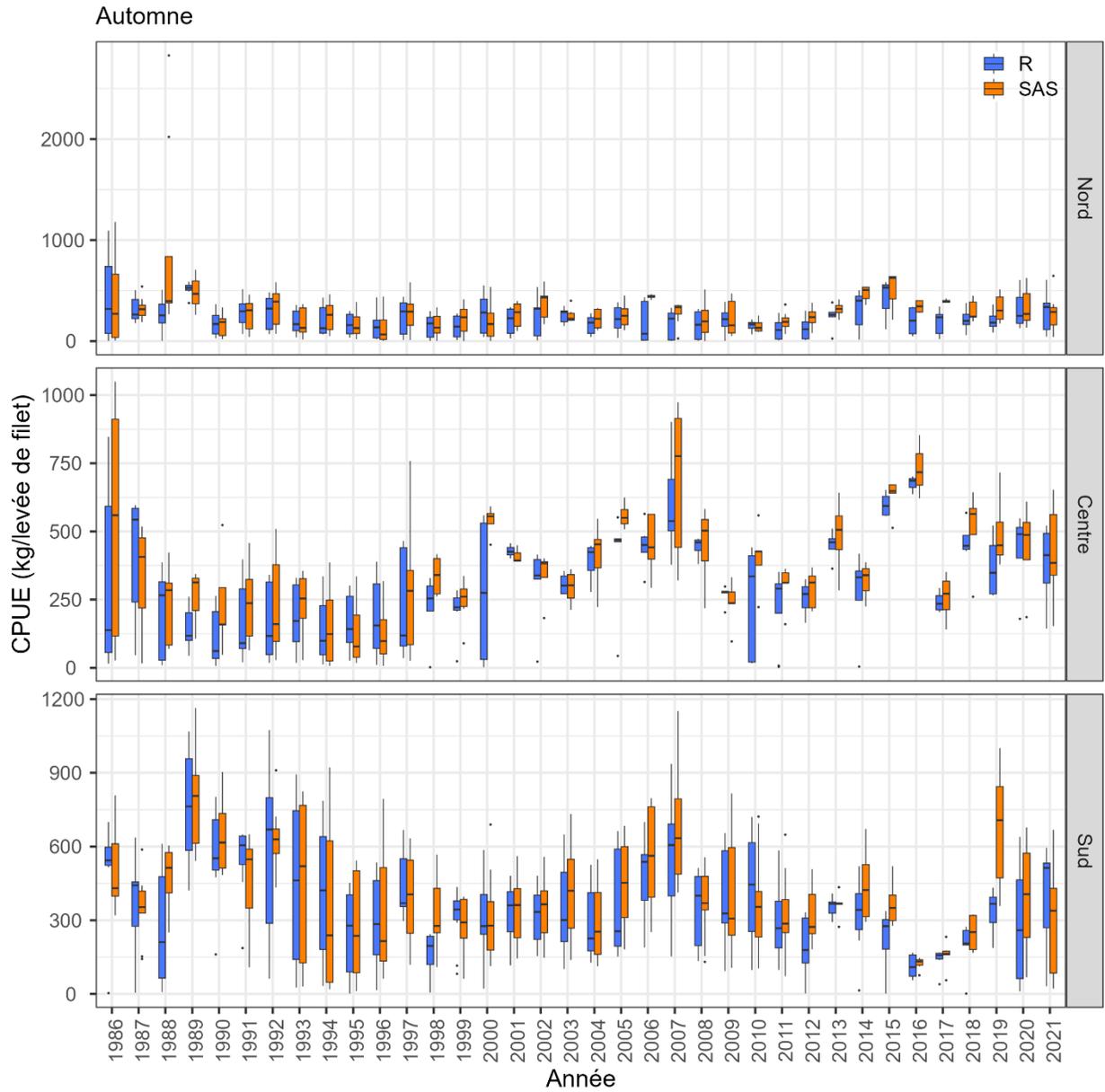


Figure 20. CPUE calculées (boîte à moustaches) par région (*large_area*: Nord, Centre, Sud) et par semaine pour la pêche automnale entre 1986 et 2021. R en bleu. SAS en orange.

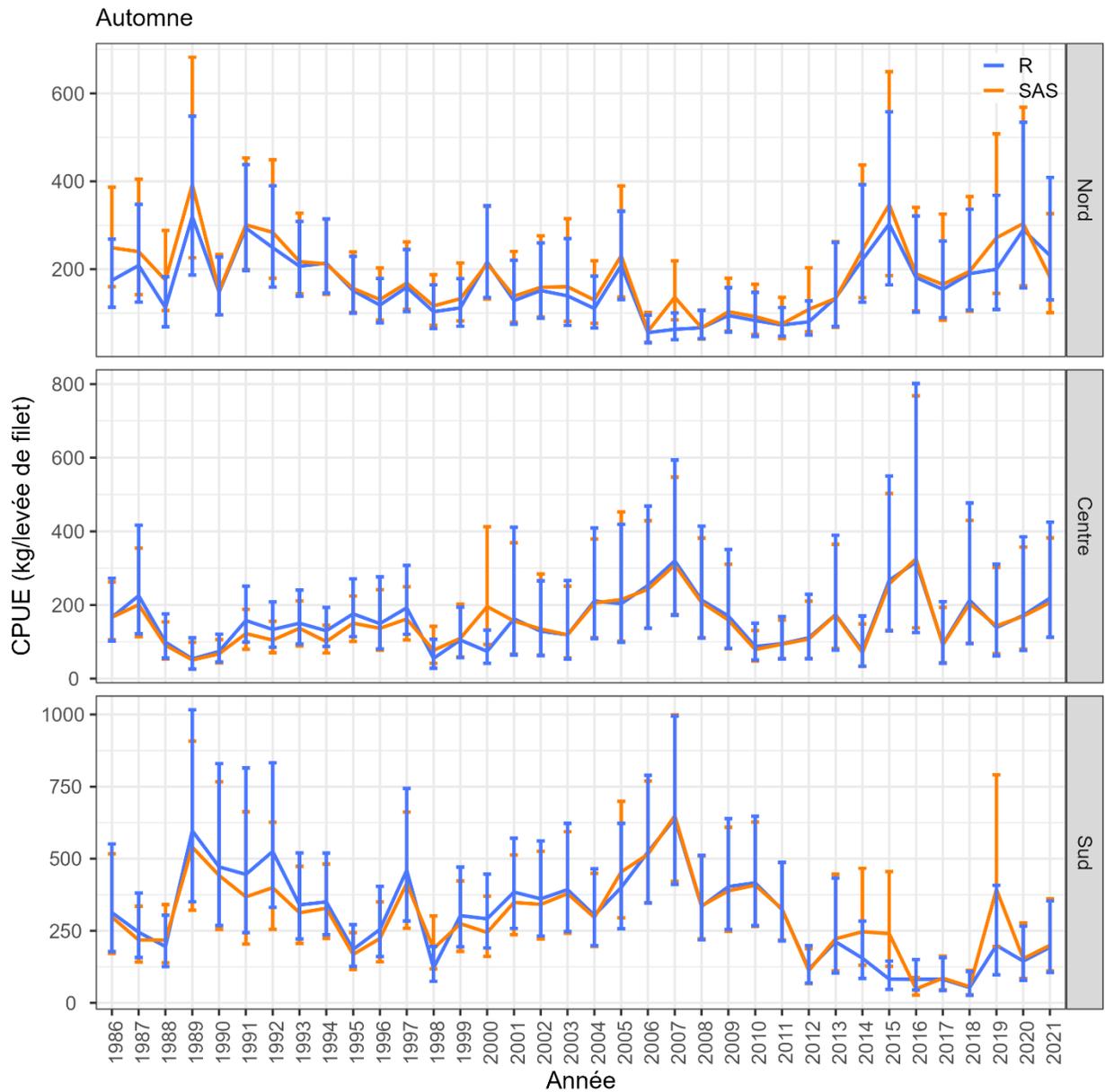


Figure 21. Estimations de la CPUE (kg par levée de filet, moyenne et intervalles de confiance 5 et 95 %) par région (*large_area*: Nord, Centre, Sud) pour la pêche automnale entre 1986 à 2021. R en bleu. SAS en orange.

Pour la pêche automnale, on remarque à la Figure 20 que la procédure SAS semble donner des valeurs plus élevées ayant une variabilité plus petite particulièrement pour la région Nord de 1988, 2006, 2007 et 2017, pour la région Centre entre 2000 et 2010 et pour la région Sud en 1988; ainsi que des valeurs plus grandes ayant une variabilité plus élevée pour la région Sud en 1998 et 2019. La valeur pour la région du Sud de 2019 est probablement explicable par la plus grande complétion des valeurs de l'effort de la procédure R, mais les autres différences sont probablement dues à plusieurs facteurs.

DISCUSSION

L'objectif de ce travail était de traduire les codes utilisés pour estimer la capture par unité d'effort (CPUE) du langage SAS vers le langage R dans le but d'améliorer l'accessibilité, la compréhension et la transférabilité des scripts.

En plus d'atteindre cet objectif, ce travail a permis :

- D'automatiser et d'uniformiser les procédures reliées à l'importation et à la sélection des données ainsi que celles reliées aux calculs effectués. Cette étape permettra d'améliorer la répliquabilité et la reproductibilité des estimations utilisées dans le modèle de population (Rolland et al. 2022).
- L'utilisation des données brutes a permis d'identifier les modifications méthodologiques apportées au cours du temps et en particulier, de compléter certaines trouées dans le calcul du nombre moyen de filets standards utilisés pour le calcul de l'effort (voir Figure 12 à Figure 14).
- De revoir et d'uniformiser les définitions des différents niveaux d'agrégation spatiales (voir Figure 4). Cette mise à jour permet d'obtenir des valeurs de CPUE calculées au plus petit niveau d'agrégation possible et permettra d'explorer de nouveaux modèles d'estimation de la CPUE.
- Il a aussi permis de réduire la redondance afin de minimiser les erreurs et d'améliorer l'adéquation de la procédure utilisée avec celle présentée dans les évaluations des sous-populations de hareng du sGSL (Rolland et al. 2022).

Cependant, certaines différences dans les résultats obtenus à chacune des étapes peuvent être observées. Certaines sont explicables, d'autres demanderont une investigation pour les comprendre et améliorer le processus. Certaines explications sont décrites ci-dessous et certaines améliorations sont proposées.

Agrégations spatiales et aires de pêches

Deux problématiques principales ont été observées pour les agrégations spatiales et aires de pêches lors de la traduction de SAS vers R.

La première concerne l'utilisation systématique d'une agrégation spatiale, soit les ports de débarquement *port_landed pour R*, à la place de l'aire de pêche et de gestion du hareng *herring_area*. Les ports de débarquements étaient utilisés par le programme SAS pour les valeurs manquantes de l'aire de pêche puisque la pêche aux engins fixes aux harengs du sGSL est une pêche côtière. Cette utilisation par la procédure R permet d'augmenter le nombre d'observations utilisable pour la modélisation de la CPUE comparativement à la procédure SAS. Cependant, on ne peut pas être certain qu'un port de débarquement correspond réellement à une aire de pêche lorsque le port est limitrophe à plus d'une aire. Une analyse pourrait être effectuée pour vérifier le pourcentage adéquation entre un port de débarquement à l'aire de pêche. De plus, les agrégations spatiales du relevé téléphonique sont définies en fonction de la réponse à la question : « Où pêchez-vous? », et non par le port de débarquement comme c'est le cas pour le relevé de monitoring à quai dans la procédure R actuelle.

La deuxième problématique est qu'aucun calcul n'est effectué pour le 4Vn dans la procédure SAS. Une partie du travail pour permettre à la procédure R de le faire a été effectué, mais les résultats ne peuvent être produits pour le moment.

Mise à jour des scripts de données sources

Un effort a été effectué pour corriger les erreurs de saisies dans les données de monitoring à quai (DMP) et du relevé téléphonique (*Telsurvey*), mais cet effort n'est pas exhaustif. Une méthodologie devrait être mise sur pied pour le faire de façon systématique dans le futur.

Une attention particulière a été portée l'utilisation des données brutes et cette décision a entraînée l'arrêt de l'utilisation des données transformées et archivées annuellement par la procédure SAS. Cependant, il est possible de les importer et de les utiliser en effectuant quelques corrections mineures dans les scripts du projet R (voir l'Appendice 1).

DMP

Il y avait une grande variabilité dans la mise en forme des données de DMP entre les années. La fonction *std_names_fct* a été créée dans le but d'uniformiser le nom des variables et leurs classes, principalement pour les données du DMP. Elle n'est pas parfaite et son fonctionnement devra être revu pour intégrer plus adéquatement les nouvelles variables du journal de bord obligatoire et ajuster les définitions et la hiérarchie des variables (c.-à-d. : variable parent et enfant). De plus, à l'aide de cette fonction d'uniformisation, les données brutes en provenance des départements des statistiques du Golfe et du Québec pourraient être importées directement sans transformation. Cette étape a été testée, mais n'est pas encore complètement implémentée.

Telsurvey

La procédure portant sur le relevé téléphonique a été revue et une partie des erreurs de saisie dans les données ont été corrigées. Ces étapes ne semblent pas avoir causé de différences dans les résultats obtenus. Ces corrections ne sont pas présentées dans ce rapport, mais ont néanmoins mis en lumière la disponibilité potentielle d'autres variables qui pourraient être utilisées pour améliorer la composante de l'effort dans le calcul de la CPUE.

ZIFF

La procédure de sélection des données de capture utilisée pour les ZIFF avant 2002 par le programme R est basée sur les informations recueillies dans les différents rapports de recherche. Bien qu'elle permette d'utiliser les données brutes plutôt que les données archivées d'année en année, elle est systématique et semble adéquate. Toutefois, des différences entre les valeurs de ZIFF sont observées entre la procédure SAS et R. En particulier, la différence pourrait être due au problème de la classe de la variable du numéro de bateau. Une autre des causes de la différence observée pour le printemps provient probablement de la complétion dans R des données de ZIFF à l'aide des bordereaux d'achat (pslips) pour les données antérieures à 2003. Une recherche

plus approfondie devrait être entreprise en collaboration avec les équipes de statistique des pêches de la région du Golfe et du Québec pour déterminer la procédure de sélection et de vérification des données avant de les inclure dans les ZIFF pour cette période.

La valeur du nombre de filets standards

Il existe des différences de sélections de la valeur du nombre de filets standard entre les programmes R et SAS. Proposer une nouvelle façon de faire a permis d'améliorer la compréhension du programme par l'utilisateur, mais aussi de correspondre à celle présentée dans les rapports de recherches. Les points ci-dessous présentent les problématiques soulevées par l'amélioration de la procédure et suggèrent des façons de les résoudre.

- Pour l'automne, bien que la procédure de choix du plus grand nombre d'observations existe, les scripts SAS et R privilégient le nombre de filets issus du relevé téléphonique parce que le nombre de levées de filet (*hauls*) y est documenté et que ce n'est pas le cas dans le programme de monitoring à quai. Le calcul du nombre de levées multiplié par le nombre de filets (*eff2_mean*) pourrait être fait après le choix du relevé ayant le plus d'observation plutôt qu'avant, ce qui permettrait d'être fidèle à la procédure de sélection décrite dans les documents de recherche (Rolland et al. 2022). Par ailleurs, le nombre de levées de filet (*hauls*) est inclus par définition dans la variable *nets_fished* pour le programme de monitoring à quai, mais la procédure actuelle et passée n'en tient pas compte. Cette composante devra être revue et passe par une revue de la définition et de l'uniformisation des noms des variables.
- La procédure actuelle décrite dans les documents de recherche fait un choix d'utilisation d'une source de données en fonction du nombre d'observations. Une analyse supplémentaire, voir l'Appendice 3, démontre qu'il serait probablement pertinent d'évaluer la possibilité de combiner les informations des deux relevés qui obtiennent des résultats similaires afin d'obtenir un nombre d'observations supérieur.
- Actuellement, lorsque la longueur d'un filet n'existe pas, il est établi à 14 brasses. La nouvelle procédure permettrait d'envisager l'utilisation de la moyenne annuelle par aire de pêche au hareng ou par agrégation spatiale de l'aire de pêche au hareng. La valeur de la longueur d'un filet permettant l'uniformisation entre les aires et les années pourrait aussi être revue chaque année.
- À l'intérieur de la fonction *fix_net_to_fathoms_fct* permettant de corriger la longueur d'un filet, les valeurs entre 50 et 56 sont assignées à la valeur 16.6 brasses. La raison pour laquelle on assigne ces valeurs de cette façon devrait être examinée, car elle est utilisée sur pratiquement toute la série temporelle (1986 à 2017).

De façon générale, il y a aussi actuellement un problème d'uniformisation du nom des variables entre les sources de données qui nuit à la compréhension de la procédure, ce problème devra être abordé en même temps que celui présenté à la section du programme de monitoring à quai (DMP).

Sélection des données de captures pour le printemps

La procédure de sélection des données de captures a été revue et élaborée pour permettre d'utiliser une seule procédure pour le printemps et pour l'automne, ce qui n'était pas le cas précédemment. Une des différences principales entre le printemps et l'automne dans la procédure SAS est la sélection des données du programme de monitoring à quai (DMP) plutôt que celle des ZIFF pour la pêche printanière, car il n'est pas clairement démontré dans les documents de recherche qui ont établi cette procédure (Poirier 2002). L'Appendice 4 se penche sur cette question et démontre qu'il était adéquat de les utiliser pour le printemps seulement, mais qu'il n'est plus requis de le faire depuis 1998.

Révision du calcul de la CPUE : le nombre d'observation

Aucune modification n'a été effectuée dans la méthode de modélisation de la CPUE choisie pour chacune des sous-populations soit à l'utilisation des modèles de régressions présentés aux équations

$$(1) \ln CPUE_{ijk} = \alpha + \beta_1 I + \beta_2 J + \beta_3 K + \varepsilon$$

$$(2) \ln CPUE_{ik} = \alpha + \beta_1 I + \beta_3 K + \varepsilon$$

(2. Il serait probablement avisé de les revoir à l'aide de nouveaux outils, surtout dans l'éventualité que les journaux de bord permettent d'obtenir des valeurs de capture égale à zéro.

La différence principale est le nombre d'observations qui entre dans le modèle. Contrairement à la procédure SAS, les données de l'effort pour la procédure R sont disponibles pour toutes les années et toutes les agrégations spatiales. Il y a aussi un plus grand nombre d'observations disponible à la suite de l'utilisation de la procédure R et c'est ce qui semble être l'explication principale à la plupart des différences observées entre les procédures dans ce rapport.

Actuellement, la procédure retire des observations pour respecter la Loi sur la protection des renseignements personnels (L.R.C. (1985), ch. P-21; «*Privacy Act*»). Elle spécifie que les pêcheurs ne doivent pas être identifiés lors de la présentation des résultats. Ainsi, toutes les lignes dont le nombre de voyages (*trips*) est inférieur ou égal à 5 sont retirées avant la modélisation. L'arrêt du retrait de ces observations permettrait d'ajouter plusieurs observations à l'entrée des modèles et ne contreviendrait probablement pas à l'esprit de la loi. Cette modification pour la procédure R permet d'augmenter le nombre d'observations de 3 077 à 6 847 pour le printemps et de 1 042 à 1 350 pour l'automne.

CONCLUSION

Les scripts présentés ont été écrits par et pour des biologistes. Ils sont en constante évolution due aux interactions avec les pairs, à l'amélioration des compétences et des connaissances des utilisateurs et des nouvelles technologies et façons de faire. Par exemple, les données devraient être importées des bases de données ou fichiers

sources directement et sans modification, un outil de correction et de vérification automatique ou semi-automatique des données et d'uniformisation des variables avant leurs intégrations devrait être créé et une procédure devrait être mise en place pour la gestion des différentes versions du projet R *herring_cpue.R* archivés sur GCode.

Pour le moment, les améliorations apportées ici n'ont pas eu d'impact sur l'évaluation des populations présentées en 2021, les résultats obtenus sont sensiblement les mêmes que ceux obtenus précédemment (Rolland et al. 2022). La procédure R est plus transparente et plus uniforme entre les années, les sources de données et les saisons de pêches. De plus, elle correspond mieux à la méthodologie décrite actuellement dans les documents de recherche. Ces commentaires et cette nouvelle procédure permettront de faciliter la mise en place de projets pour apporter des améliorations.

REMERCIEMENTS

Nous voulons remercier les réviseurs Caroline Lehoux, Kim Émond et Jacob Burbank le co-éditeur pour les Publications secondaires du MPO, pour leurs commentaires constructifs et leurs apports à l'amélioration de ce manuscrit.

RÉFÉRENCES

- Bivand, R., Keitt, T., et Rowlingson, B. 2021. rgdal: Bindings for the “geospatial” data abstraction library. Available from <https://CRAN.R-project.org/package=rgdal>.
- Chadwick, E.M.P., et Cairns, D.K. 1988. Assessment of Atlantic herring in NAFO Division 4T, 1987. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **88/38**: 53 p.
- Chadwick, E.M.P., Cairns, D.K., et Dupuis, H M. 1989. Assessment of Atlantic herring in NAFO Division 4T, 1988. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **89/63**: 58 p.
- Chadwick, E.M.P., et Nielsen, G.A. 1986. Assessment of Atlantic herring in NAFO Division 4T, 1985. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **86/38**: 47 p.
- Chadwick, E.M.P., et Nielsen, G.A. 1987. Assessment of Atlantic herring in NAFO Division 4T, 1986. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **87/80**: 48 p.
- Chaput, G., Mowbray, F., Dupuis, H., Bourque, C., MacDougall, C., et Mello, L. 1993. Assessment of the NAFO Division 4T Atlantic herring stock, 1992. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **93/59**: 95 p.
- Clayton, R., Dupuis, H., Mowbray, F., Nielsen, G., Paulin, L., et Bourque, C. 1995. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stock, 1994. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **95/69**: 136 p.
- Clayton, R., LeBlanc, C., Dale, J., Nielsen, G., et Paulin, L. 1996. Assessment of the NAFO Division 4T Southern Gulf of St Lawrence Herring Stock, 1995. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **96/79**: 137 p.
- Clayton, R., LeBlanc, C., MacDougall, C., et Poirier, G. 1998. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stock, 1997. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **98/47**: 156 p.

- Claytor, R.R., Nielsen, G., Dupuis, H.M.C., et Cairns, D.K. 1991. Assessment of Atlantic herring in NAFO division 4T, 1990. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **91/66**: 76 p.
- Dowle, M., et Srinivasan, A. 2021. data.table: Extension of `data.frame`. Available from <https://CRAN.R-project.org/package=data.table>.
- Grolemund, G., et Wickham, H. 2011. Dates and times made easy with lubridate. J. Stat. Softw. **40**(3): 1–25.
- Hester, J., et Bryan, J. 2021. glue: Interpreted string literals. Available from <https://CRAN.R-project.org/package=glue>.
- Horikoshi, M., et Tang, Y. 2016. ggfortify: Data visualization tools for statistical analysis results. Available from <https://CRAN.R-project.org/package=ggfortify>.
- Lamichhaney, S., Fuentes-Pardo, A.P., Rafati, N., Ryman, N., McCracken, G.R., Bourne, C., Singh, R., Ruzzante, D.E., et Andersson, L. 2017. Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. Proc. Natl. Acad. Sci. **114**(17): E3452–E3461. doi:10.1073/pnas.1617728114.
- LeBlanc, C., et LeBlanc, L. 1996. The 1995 NAFO Division 4T Herring Gillnet Telephone Survey. Can. Atl. Fish. Sci. Advis. Comm. Res. Doc. **96/77**: 37 p.
- LeBlanc, C.H., MacDougall, C., et Bourque, C. 2008. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stocks in 2007. Can. Sci. Advis. Secr. Res. Doc. **2008/061**: iv + 133 p.
- LeBlanc, C.H., MacDougall, C., Bourque, C., Morin, R., et Swain, D. 2009. Évaluation des stocks de hareng de la zone 4T de l'OPANO dans le sud du golfe du Saint-Laurent en 2008. Can. Sci. Advis. Secr. Res. Doc. **2009/049**: iv + 175 p.
- LeBlanc, C.H., MacDougall, C., et Poirier, G.A. 2002. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stocks in 2001. Can. Sci. Advis. Secr. Res. Doc. **2002/053**: 102 p.
- LeBlanc, C.H., Mallet, A., MacDougall, C., Bourque, C., et Swain, D. 2012. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stocks in 2011. Can. Sci. Advis. Secr. Res. Doc. **2012/111**: vi + 167 p.
- LeBlanc, C.H., Poirier, G.A., MacDougall, C., Bourque, C., et Roy, J. 2007. Assessment of the NAFO Division 4T southern Gulf of St. Lawrence herring stocks in 2006. Can. Sci. Advis. Secr. Res. Doc. **2007/016**: i + 113 p.
- Lenth, R.V. 2022. emmeans: Estimated marginal means, aka least-squares means. Available from <https://CRAN.R-project.org/package=emmeans>.
- McDermid, J.L., Mallet, A., et Surette, T. 2016. Fishery performance and status indicators for the assessment of the NAFO Division 4T southern Gulf of St. Lawrence Atlantic herring (*Clupea harengus*) to 2014 and 2015. Can. Sci. Advis. Secr. Res. Doc. **2016/060**: ix + 62 p.
- Ooms, J. 2021. magick: Advanced graphics and image-processing in r. Available from <https://CRAN.R-project.org/package=magick>.
- Oracle. 2017. Oracle Database Online Documentation. Release 1 (12.1).
- Pebesma, E.J., et Bivand, R.S. 2005. sp: Classes and methods for spatial data in R. Available from <https://cran.r-project.org/web/packages/sp/index.html>.

- Poirier, G.A. 2002. Catch and effort data used in the assessment of herring populations in the southern Gulf of St Lawrence. Can. Sci. Advis. Secr. Res. Doc. **2002/61**: 28 p.
- Quintans, D. 2021. librarian: Install, update, load packages from CRAN, “GitHub”, and “bioconductor” in one step. Available from <https://CRAN.R-project.org/package=librarian>.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from <https://www.R-project.org/>.
- Ripley, B., et Lapsley, M. 2021. RODBC: ODBC database access. Available from <https://CRAN.R-project.org/package=RODBC>.
- Rolland, N., Turcotte, F., McDermid, J.L., DeJong, R.A., et Landry, L. 2022. Assessment of the NAFO Division 4TVn southern Gulf of St. Lawrence Atlantic herring (*Clupea harengus*) in 2020-2021. Can. Sci. Advis. Secr. Res. Doc. **2022/068**: xii + 142 p.
- SAS Institute Inc. 2013. Reference. SAS Institute Inc., Cary, NC.
- Stobo, W.T. 1982. Tagging studies on Scotian Shelf herring. Northwest Atl. Fish. Organ. Sci. Counc. Res. Document **82/IX/108**(N617): 16 p.
- Surette, T., Vergara, P., et Ricard, D. 2021. gulf: Access, manipulate, display and analyze southern Gulf data. manual.
- Wheeler, J.P., Squires, B., et Williams, P. 2006. An assessment of Newfoundland East and South Coast herring stocks to the Spring of 2006. Can. Sci. Advis. Secr. Res. Doc. **2006/101**: iv + 88 p.
- Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York. Available from <https://ggplot2.tidyverse.org>.
- Wickham, H. 2019. stringr: Simple, consistent wrappers for common string operations. Available from <https://CRAN.R-project.org/package=stringr>.
- Wickham, H. 2021. tidyr: Tidy messy data. Available from <https://CRAN.R-project.org/package=tidyr>.
- Wickham, H., et Bryan, J. 2019. readxl: Read excel files. Available from <https://CRAN.R-project.org/package=readxl>.
- Wickham, H., François, R., Henry, L., et Müller, K. 2021a. dplyr: A grammar of data manipulation. Available from <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H., Hester, J., et Bryan, J. 2021b. readr: Read rectangular text data. Available from <https://CRAN.R-project.org/package=readr>.
- Zhu, H. 2021. kableExtra: Construct complex table with “kable” and pipe syntax. Available from <https://CRAN.R-project.org/package=kableExtra>.

APPENDICE 1

La version 9.4 (SAS Institute Inc. 2013) est utilisée pour produire les résultats qui sont comparés avec ceux obtenus par les programmes R. L'estimation de la CPUE décrite dans le document : *4T HERRING GILLNET CPUE instructions.docx*, permet d'exécuter manuellement une série de programmes écrits dans le langage SAS. La procédure SAS comprend 7 scripts principaux. Le flux d'exécution des scripts listés ci-dessous est représenté à la Figure 22 et à la Figure 23.

- *2-reform_dmp_ll.sas*, importe et fusionne les données du DMP sur les captures et le nombre de filets utilisés pour les régions du Québec et du Golfe. Il est exécuté annuellement pour les saisons d'automne et de printemps. Les données sont ensuite archivées par année et par saison.
- *3-merge_dmp_net_lands_data_ll.sas*, importe et fusionne les données du DMP sur les captures et le nombre et la longueur des filets utilisés de la région du Golfe. Il est exécuté annuellement pour les saisons d'automne et de printemps. Les données sont ensuite archivées par année et par saison.
- *6-effind_ll.sas*, importe et met en forme les données brutes sur l'effort du relevé téléphonique annuel de 1986 à aujourd'hui. Les données archivées dans 2 fichiers selon leurs types d'agrégations, soit :
 - par année, par saison et par aire du relevé téléphonique ou,
 - par année et par saisons.
- *5-calc number of nets phone dmp_ll.sas*, calcul le nombre moyen de filets standards utilisés par année, par saison et par aire du relevé téléphonique pour les fichiers du relevé téléphonique et ceux du DMP. Il permet aussi de choisir, entre ces 2 sources de données, celui possédant le plus grand nombre d'observations du nombre de filets utilisés, par aire du relevé téléphonique par année ou par année seulement. Le nombre d'observation est le nombre de bateaux ou de responsables d'un bateau. C'est ce nombre de filets moyens qui sera utilisé dans le calcul de la CPUE. Ce script est exécuté annuellement pour l'automne, puis pour le printemps. Si le résultat du DMP est supérieur à celui du Telsurvey, il est archivé («copié-collé» manuellement) dans les scripts *9-cpue_fall_hauls_ll.sas* (pour l'automne) et *10-cpue_spring_ll.sas* (pour le printemps).
- *8-hergrade_ll.sas*, importe, sélectionne et met en forme les données brutes des ZIFF correspondant à l'année en cours. Les données sont agrégées par année, mois, jour, province et district statistique pour la capture en kg et pour le nombre de voyages effectués. Il est exécuté annuellement pour les saisons d'automne et de printemps. Les données sont ensuite archivées par année.
- *9-cpue_fall_hauls_ll.sas*, importe les données archivées précédentes pour toutes les années, calcul et de modélise la CPUE pour les géniteurs ou pêche d'automne. Il est exécuté annuellement ou en fonction des besoins de l'estimation pour l'évaluation de stock.
- *10-cpue_spring_ll.sas*, importe les données archivées précédentes pour toutes les années, calcul et de modélise la CPUE pour les géniteurs ou pêche de printemps. Il est exécuté annuellement ou en fonction des besoins de l'estimation pour l'évaluation de stock.

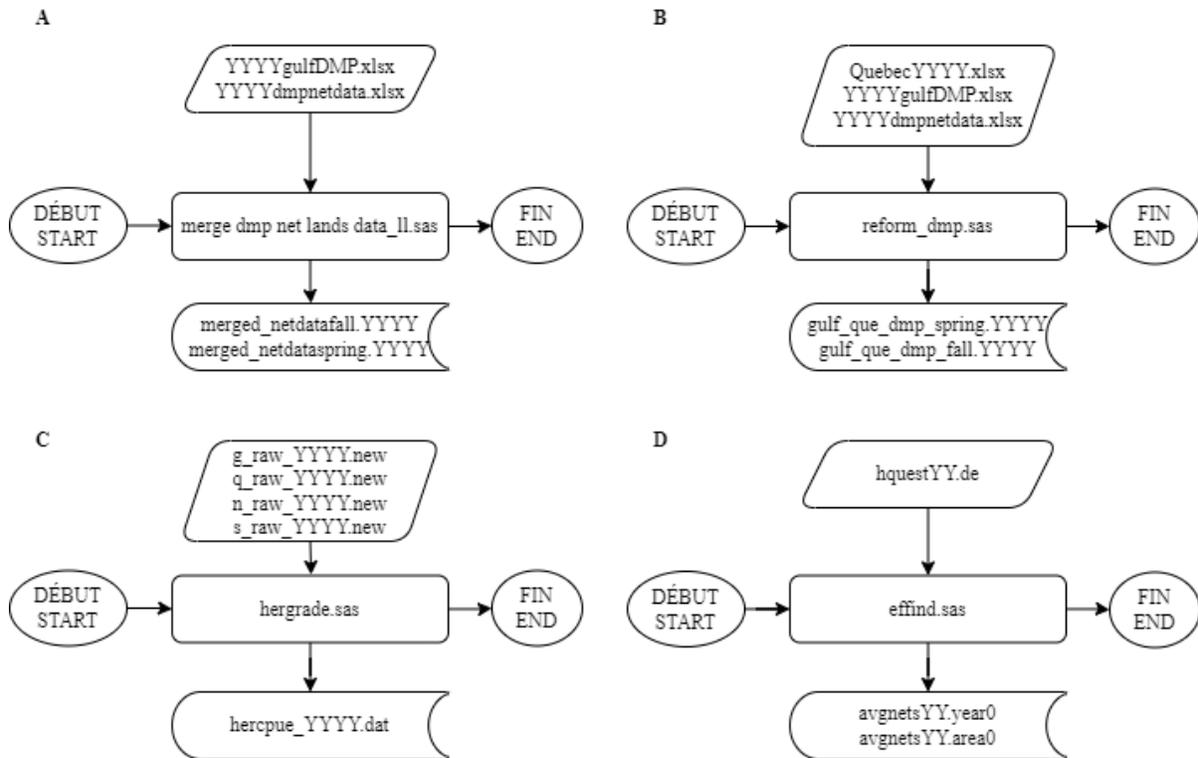


Figure 22. Diagramme du flux des opérations illustrant les scripts SAS nécessaires à l'importation et la transformation et à l'archivage des données pour la réalisation du calcul de capture par unité d'effort (CPUE). La procédure illustrée par type de données sources : A : DMP (seulement les

données du Golfe), B : DMP, C : ZIFF, D : Telsurvey, est celle qui est généralement utilisée. Certaines différences sont observées entre les années.

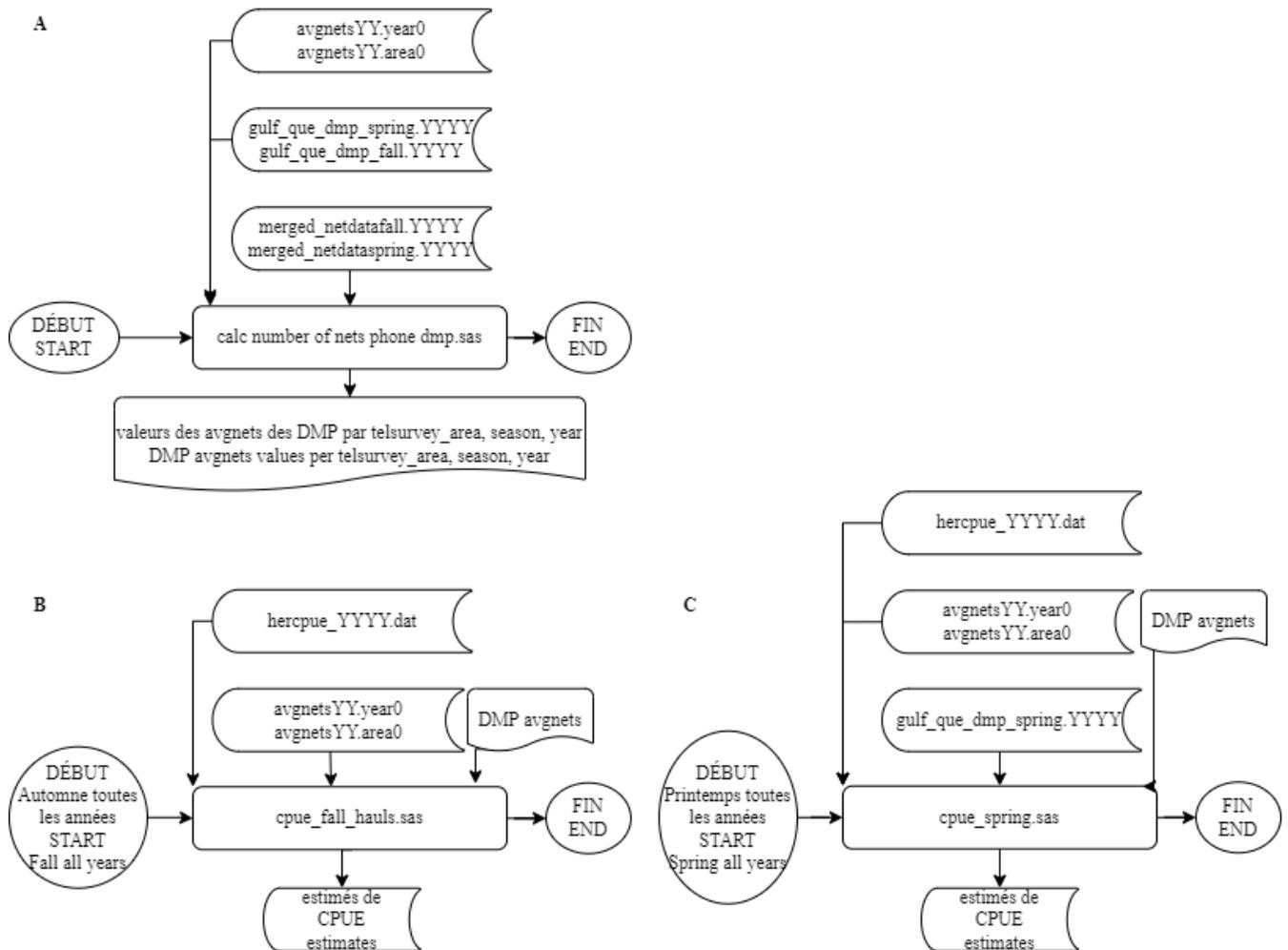


Figure 23. Diagramme du flux des opérations illustrant les scripts SAS nécessaires à l'importation des données archivées et à la réalisation du calcul de capture par unité d'effort (CPUE). La procédure illustrée : A : calcul de l'effort, B : estimation CPUE pour l'automne, C : estimation CPUE pour le printemps, est celle qui est généralement utilisée. Certaines différences sont observées entre les années.

Dans le projet *herring_cpue.R* archivés sur GCode on retrouve aussi un script où on retrouve les fonctions pour importer les données archivées par les programmes SAS, mais elles ne seront pas décrites dans ce document.

APPENDICE 2

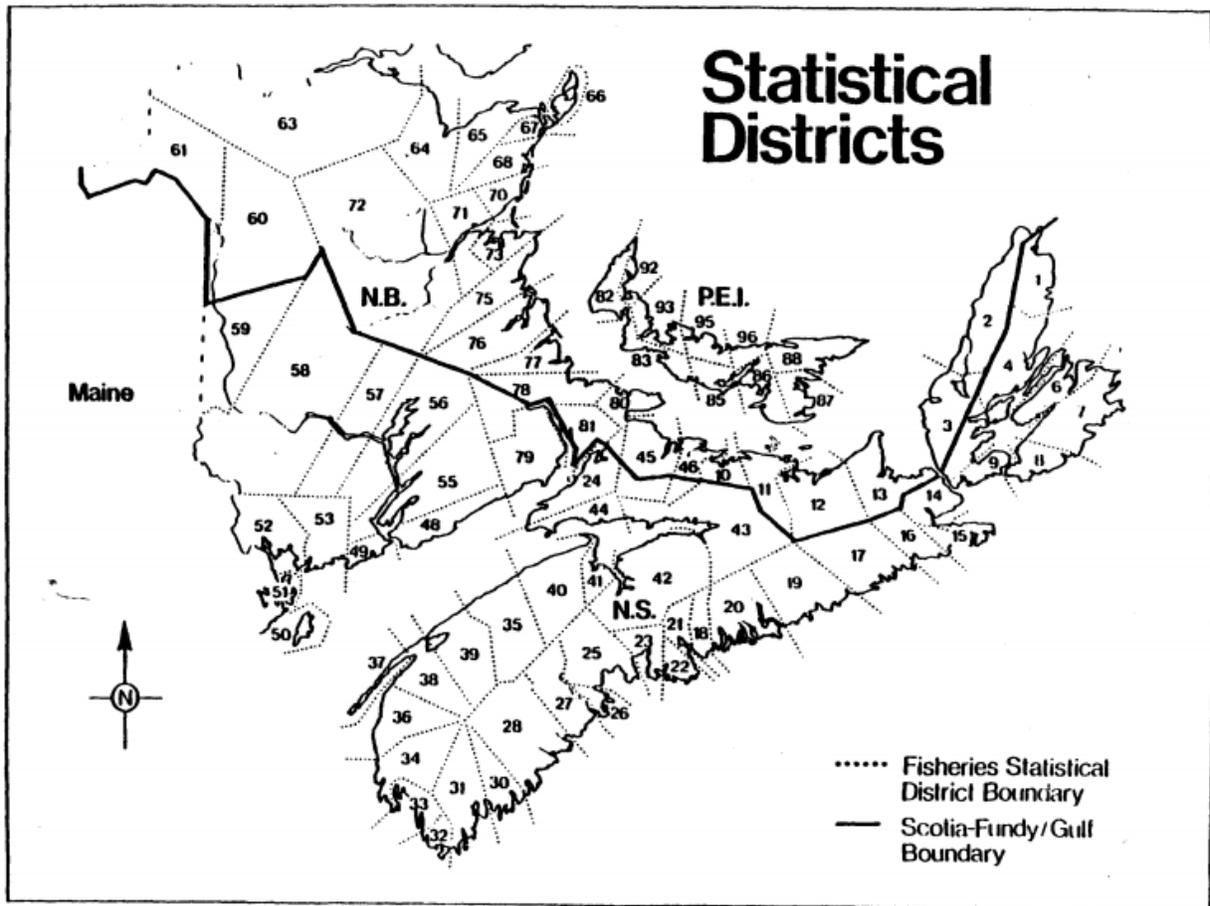


Fig. 3. Maritime Statistical Districts.
Fig 3. Districts statistiques des Maritimes.

Figure 24. Districts statistiques de l'Île-du-Prince-Edward (P.E.I.), du Nouveau-Brunswick (N.B) et de la Nouvelle-Écosse (N.S.), correspondant à la Figure 3 de Claytor et al. 1991.

APPENDICE 3

Est-ce que les résultats sont similaires entre le DMP et le Telsurvey pour le nombre moyen de filets?

Pour permettre de répondre à la première question, les Figure 25 à Figure 33 font la comparaison graphique entre les données issues du DMP et du Telsurvey pour le nombre moyen de filets standards par bateau et le nombre d'observations; par année et par aire du relevé téléphonique (1 à 8). La Figure 33 effectue cette même comparaison par année et pour l'ensemble des aires du relevé téléphonique.

On remarque que les nombres moyens de filets standards obtenus sont similaires entre les 2 relevés. On remarque aussi qu'il y a moins de valeur disponible pour le DMP que pour le Telsurvey. Pour le relevé du DMP, il n'y a aucune valeur disponible avant 1999, aucune ou peu de valeurs sont disponibles avant 2007 pour les aires 1, 3, 4 et 6, et aucune valeur n'est disponible pour l'aire 2. Cependant, dans la plupart des cas, lorsque la valeur du DMP est disponible elle possède un plus grand nombre d'observations (ou de bateaux participants) que le Telsurvey.

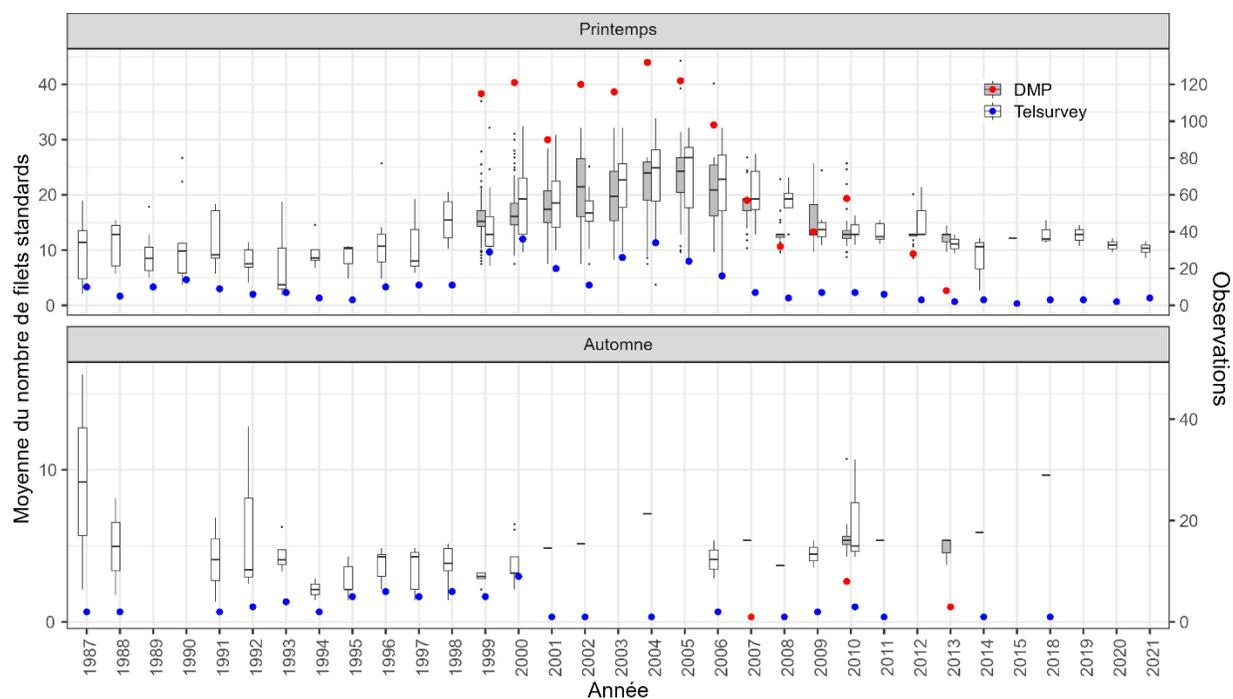


Figure 25. Aire du relevé téléphonique 1 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

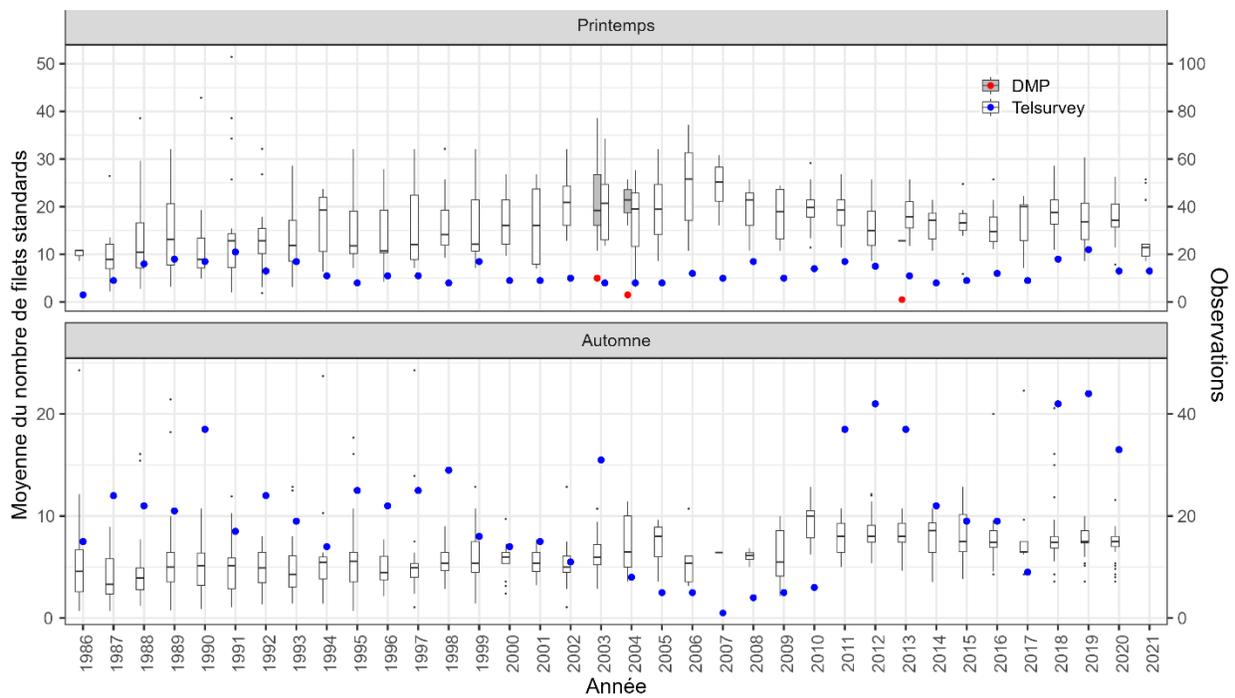


Figure 26. Aire du relevé téléphonique 2 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

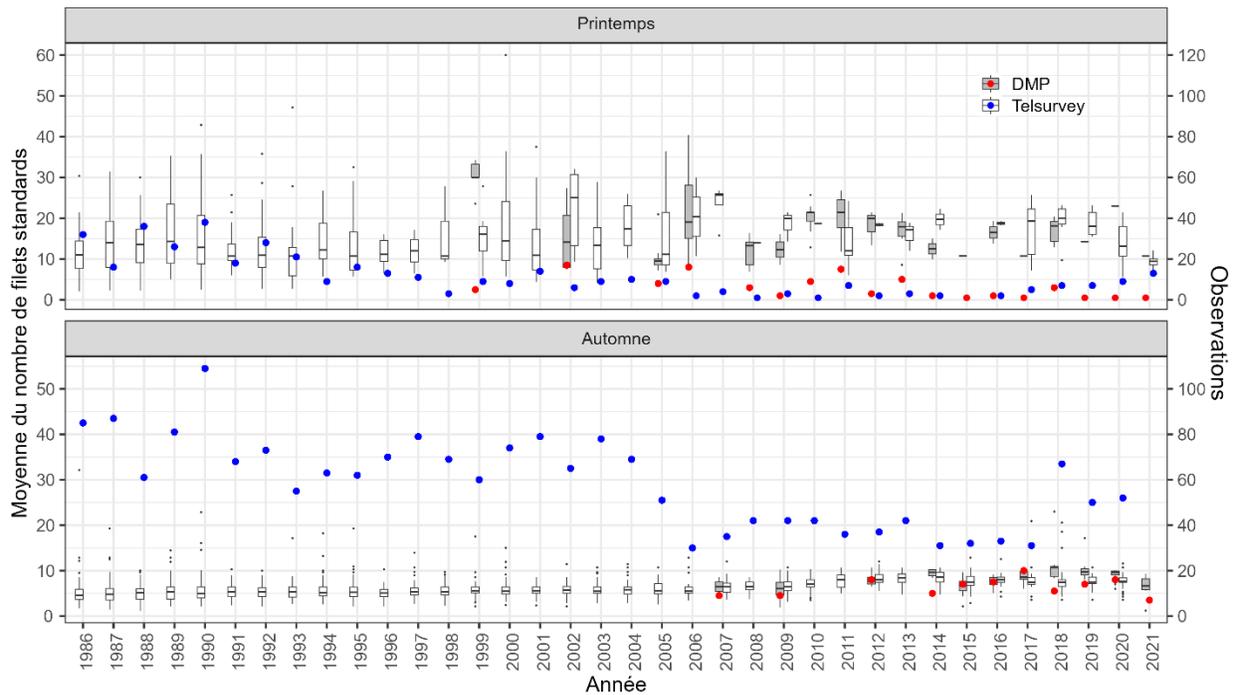


Figure 27. Aire du relevé téléphonique 3 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

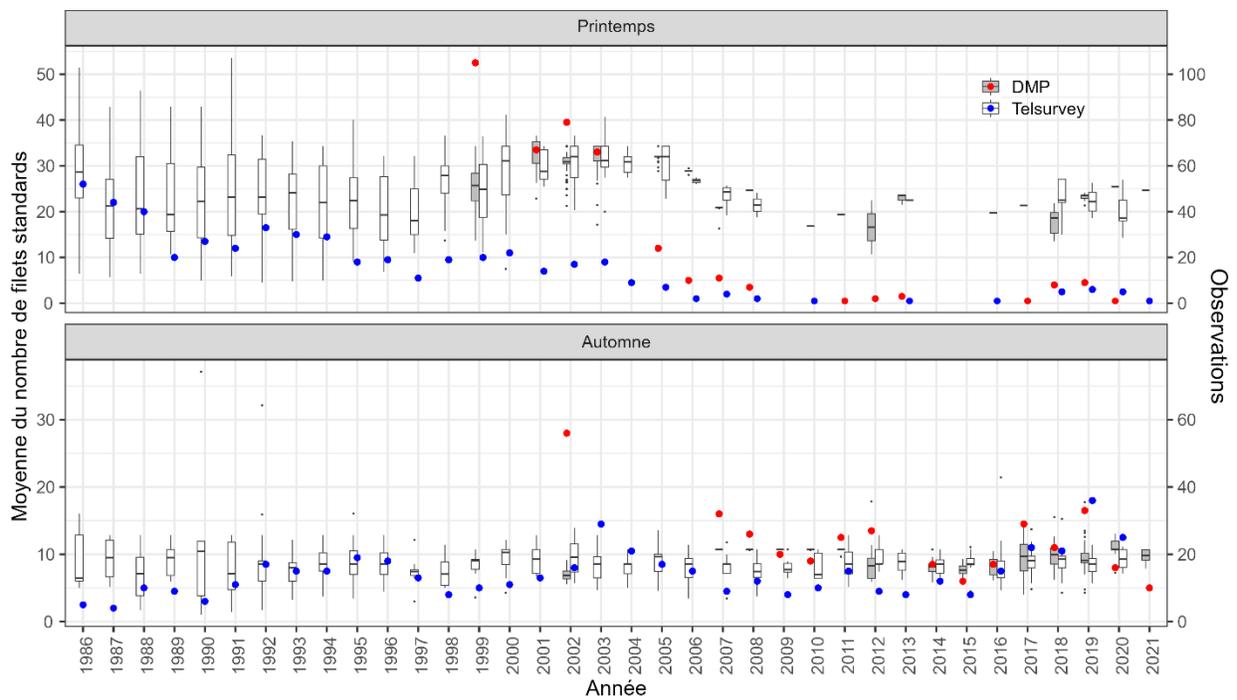


Figure 28. Aire du relevé téléphonique 4 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

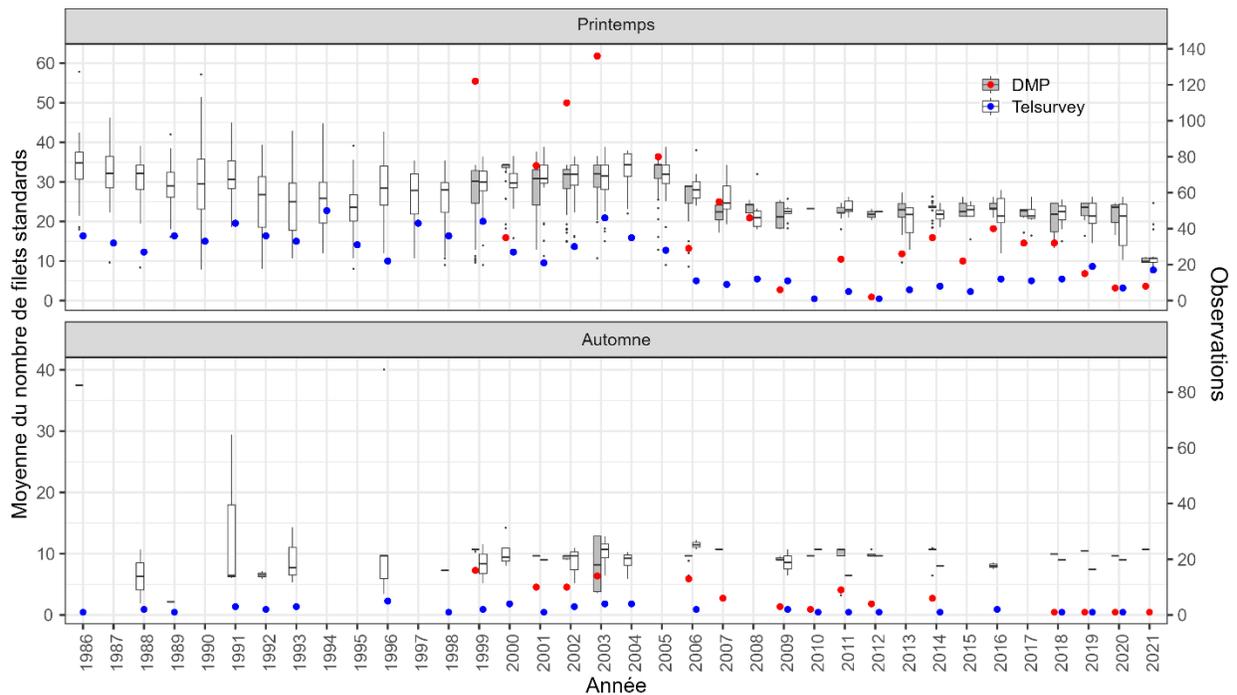


Figure 29. Aire du relevé téléphonique 5 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

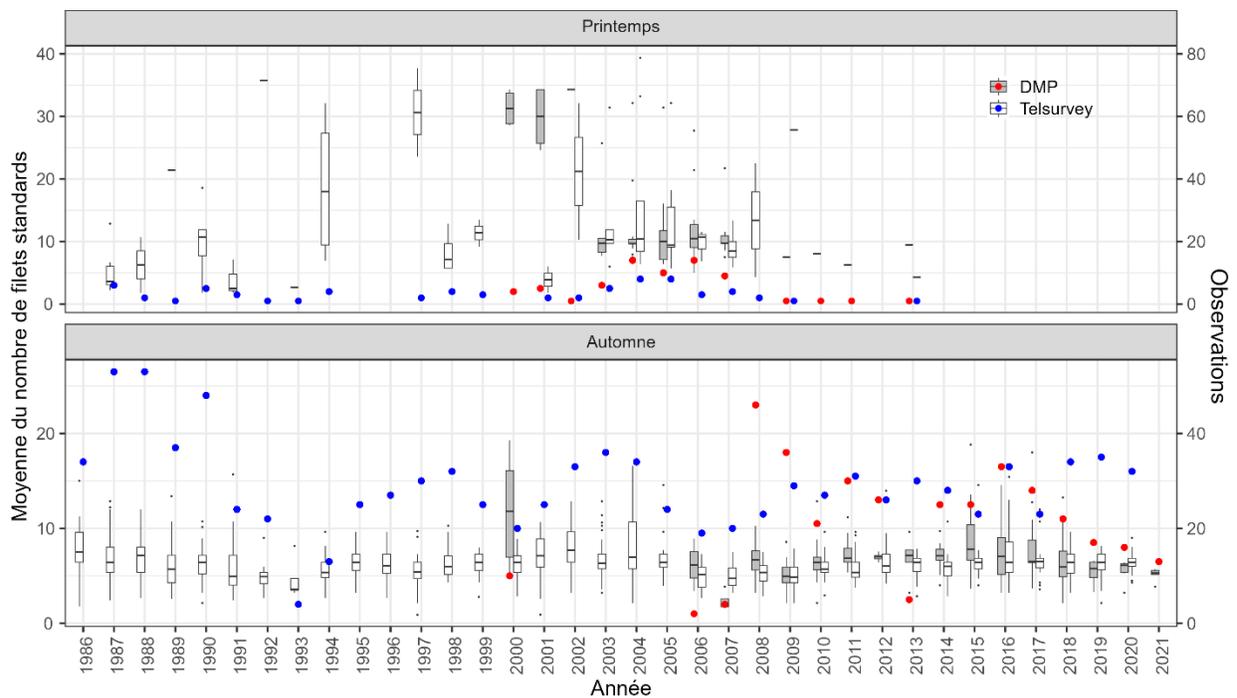


Figure 30. Aire du relevé téléphonique 6 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

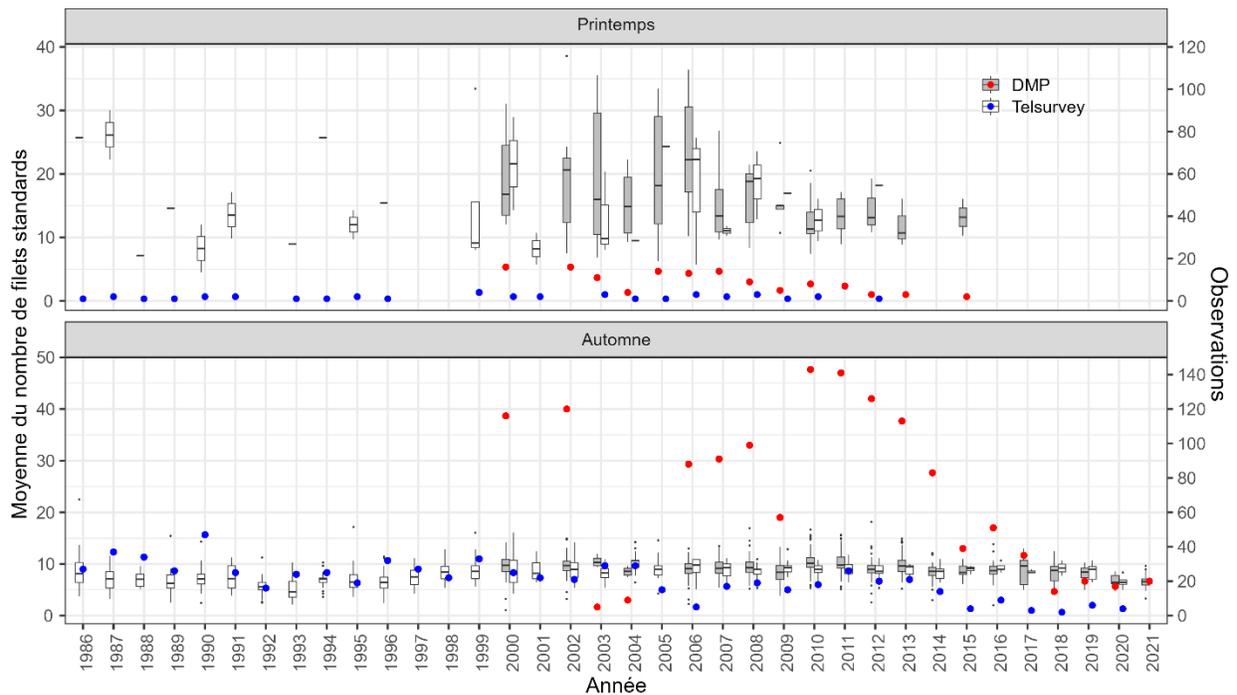


Figure 31. Aire du relevé téléphonique 7 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

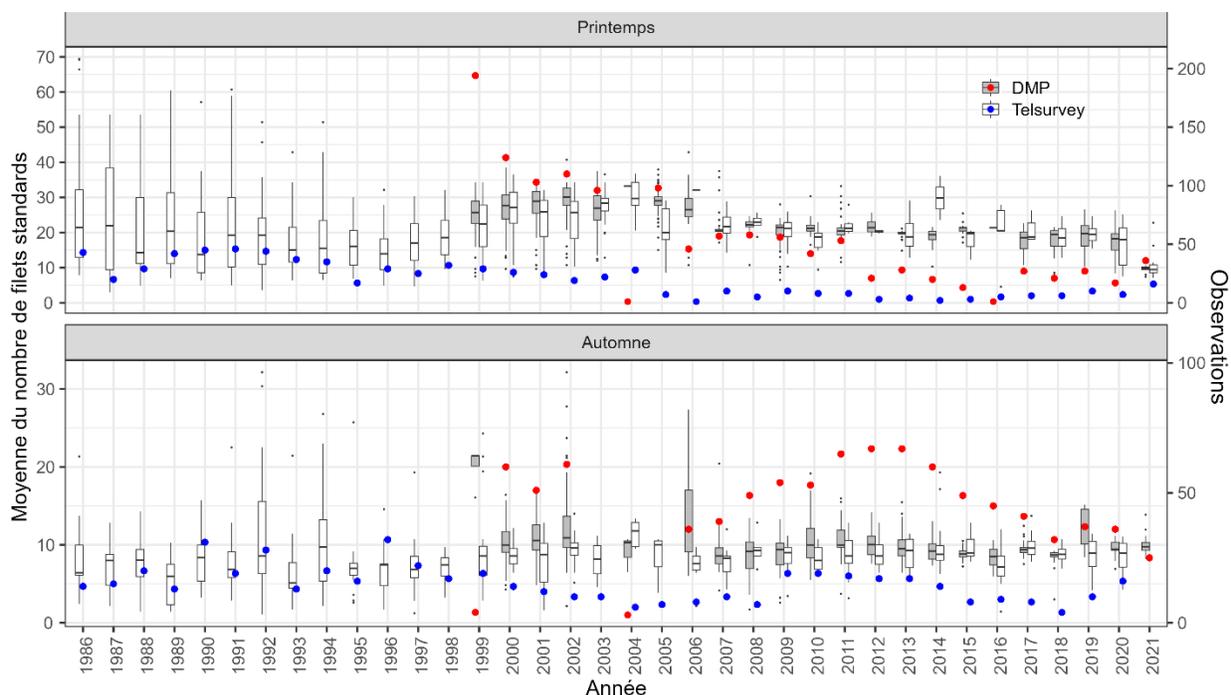


Figure 32. Aire du relevé téléphonique 8 : Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale. DMP et gris et rouge. Telsurvey en blanc et bleu.

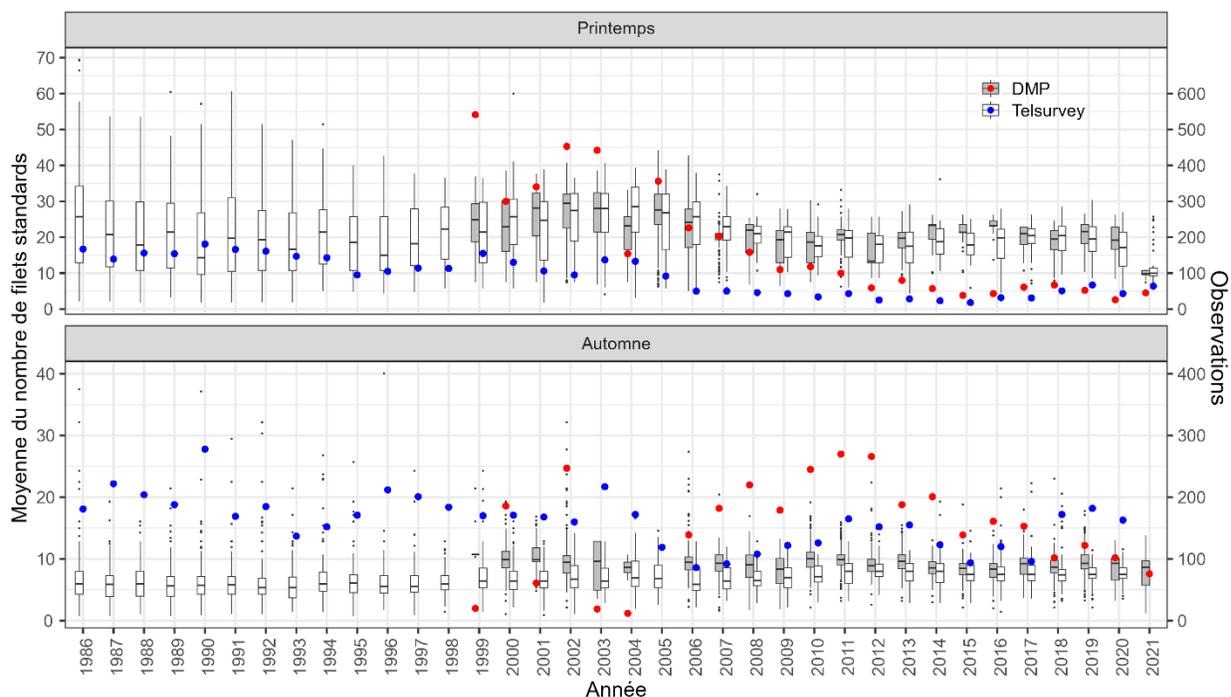


Figure 33. Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automnale pour toutes les aires du relevé téléphonique (1 à 8). DMP et gris et rouge. Telsurvey en blanc et bleu.

Serait-il préférable de combiner les 2 sources de données (DMP et Telsurvey)?

La Figure 37 montre que la combinaison des 2 sources de données permettrait d'obtenir plus d'observations (*cfv* ou *resp*), surtout pour l'automne. Cependant, il arrive qu'un même bateau pour une saison et une année en particulier ne soit pas recensé dans la même aire du relevé téléphonique entre les 2 sources de données. On observe 12 % d'erreur de ce type (147 lignes problématiques (104 bateaux) sur 1239 lignes totales) lorsqu'on fait une jointure entre les deux fichiers par année, saison, aire du relevé téléphonique, bateau. Ces erreurs pourraient entraîner des problèmes dans l'évaluation de l'effort. Les erreurs proviennent de l'identification de la zone de pêche qui est différente entre les deux relevés.

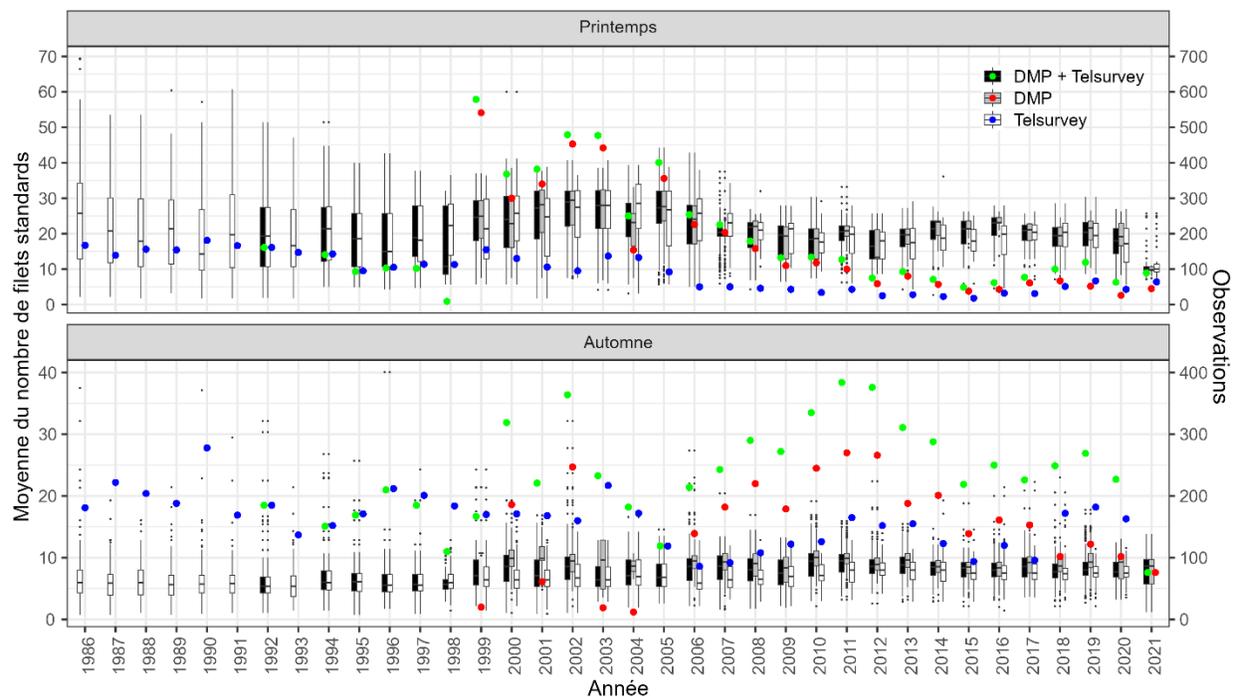


Figure 34. Nombre moyen de filets standards (boîte à moustaches) et nombre d'observations (cercles) pour la pêche printanière et l'automne pour toutes les aires du relevé téléphonique (1 à 8). DMP et gris et rouge. Telsurvey en blanc et bleu. Combinaison des deux en noir et vert.

APPENDICE 4

Représentation du décalage entre les captures pour le DMP et les ZIFF

Au début des années 90, la procédure de sélection des données de capture pour le calcul de la capture pour le printemps a été modifiée pour tenir compte du décalage qui était présent entre les données de capture du DMP et des bordereaux d'achat (Claytor et al. 1995). Cette section permet de regarder s'il est encore nécessaire aujourd'hui de privilégier les DMP pour le printemps et s'il serait pertinent pour l'automne de compléter les ZIFF avec les districts qui serait uniquement présent dans les DMP.

La Figure 35, la Figure 36 et la Figure 37 pour le printemps; ainsi que la Figure 38 et Figure 39 pour l'automne montrent les résultats des sommes de capture par semaine pour le DMP et les ZIFF. Il s'agit de la somme des captures pour les districts recensés par le DMP qui est comparé. Pour le printemps, on remarque un décalage plus important entre les DMP et les ZIFF de 1990 jusqu'à 1998. 1998 correspond à l'année où la procédure SAS arrête d'utiliser les ZIFF pour compléter les données. Pour l'automne, on ne remarque pas de décalage particulier.

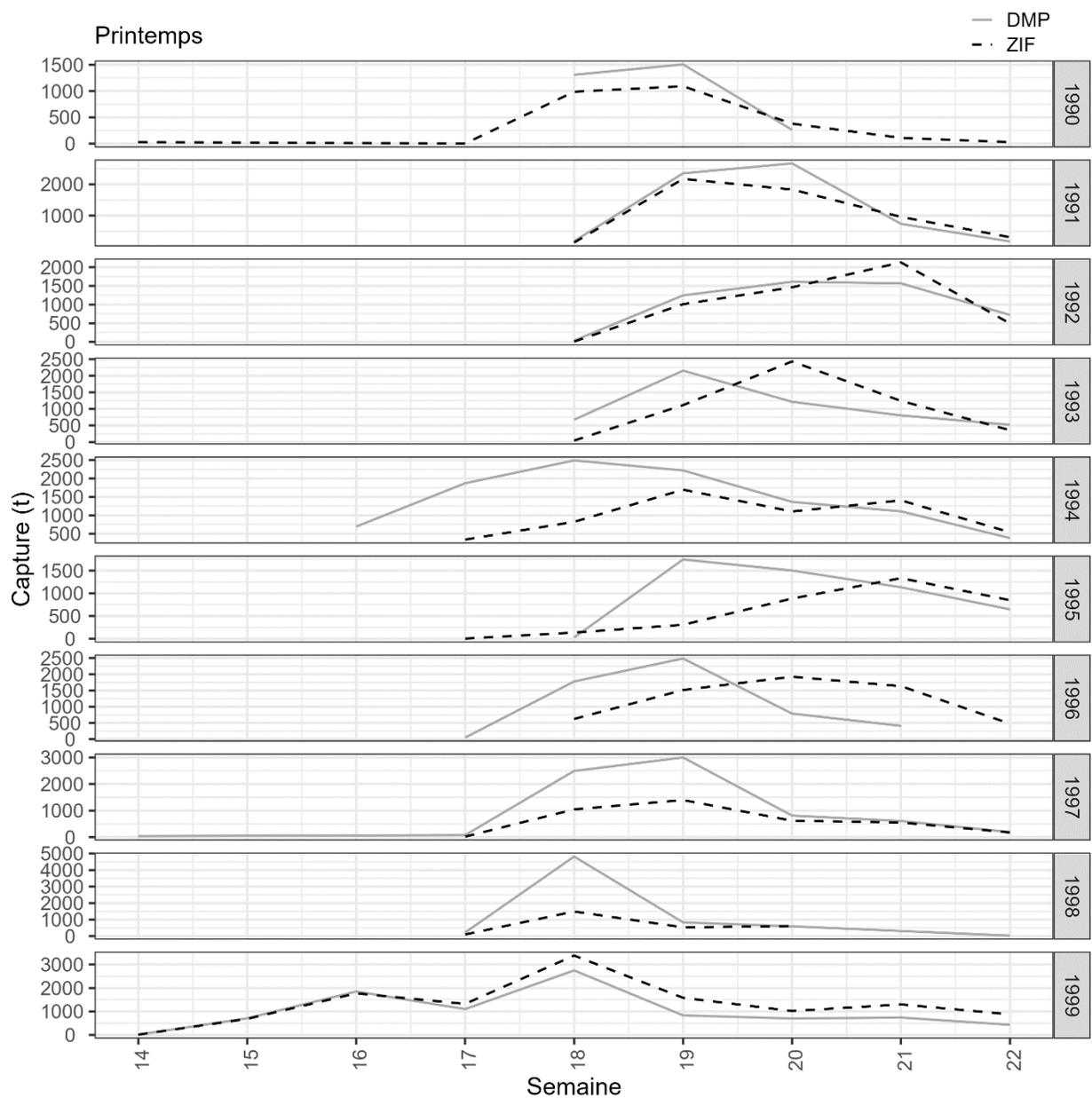


Figure 35. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 1990 et 1999. DMP en gris. ZIFF en noir hachuré.

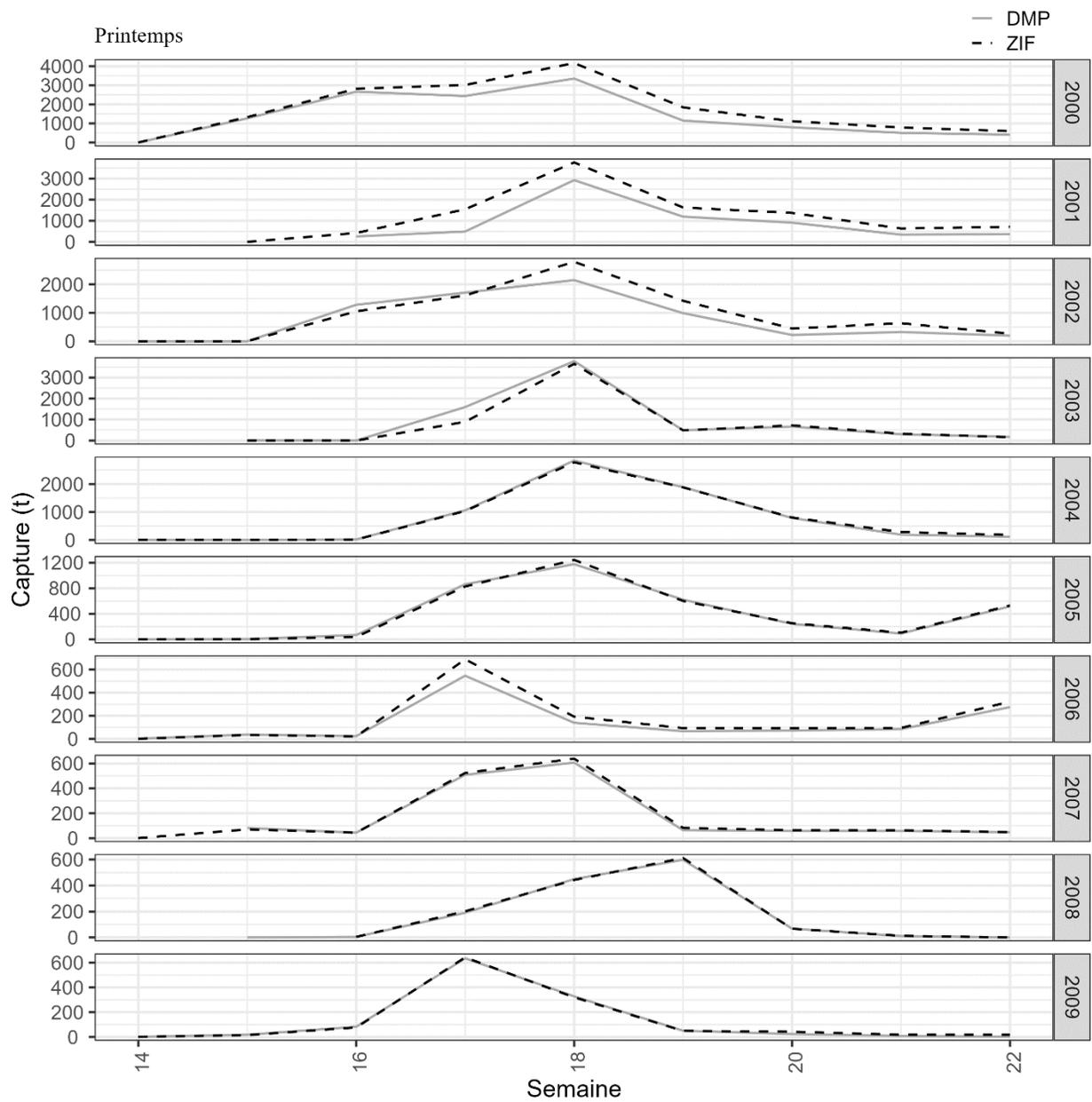


Figure 36. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 2000 et 2009. DMP en gris. ZIFF en noir hachuré.

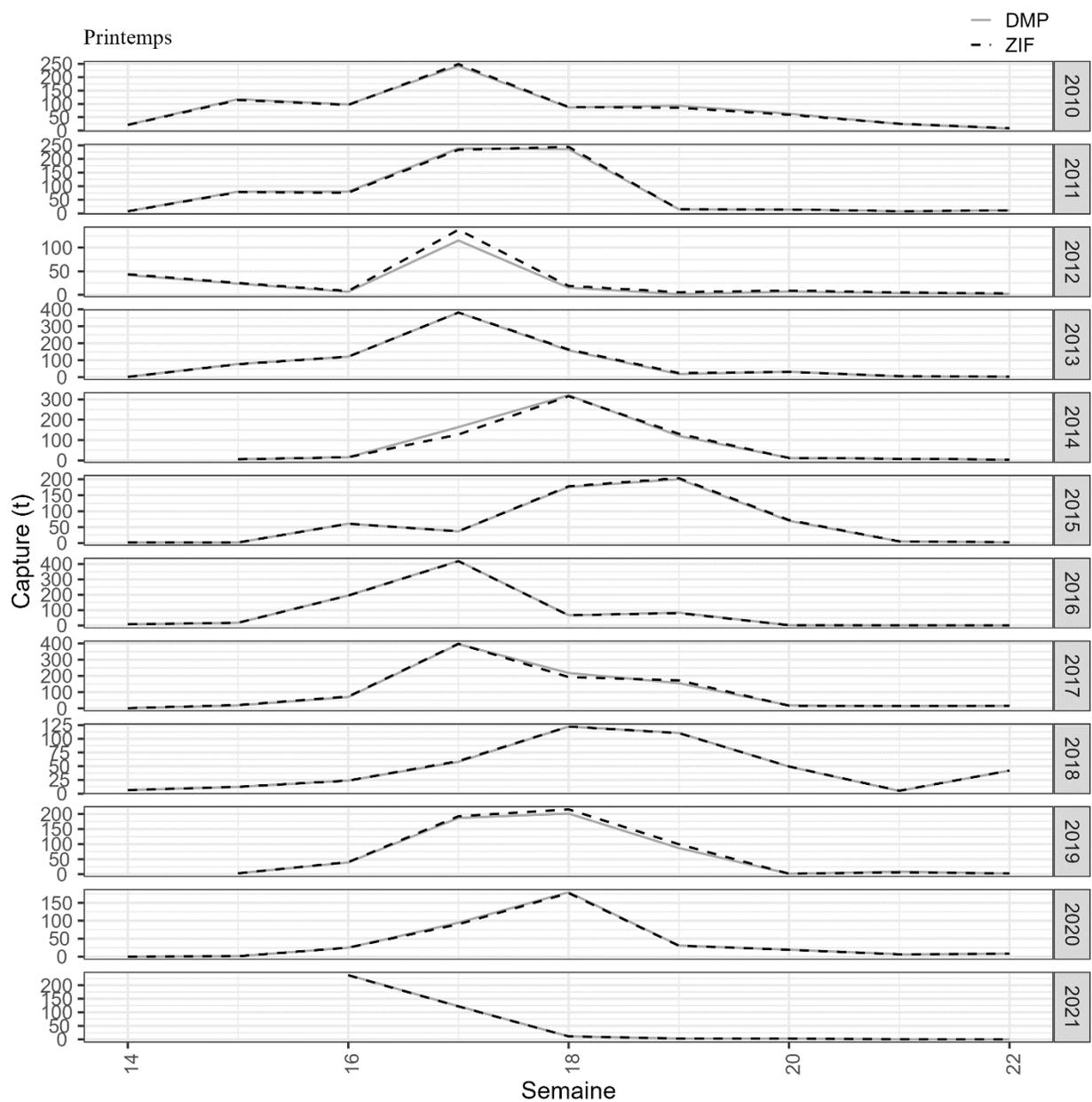


Figure 37. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche printanière entre 2010 et 2021. DMP en gris. ZIFF en noir hachuré.

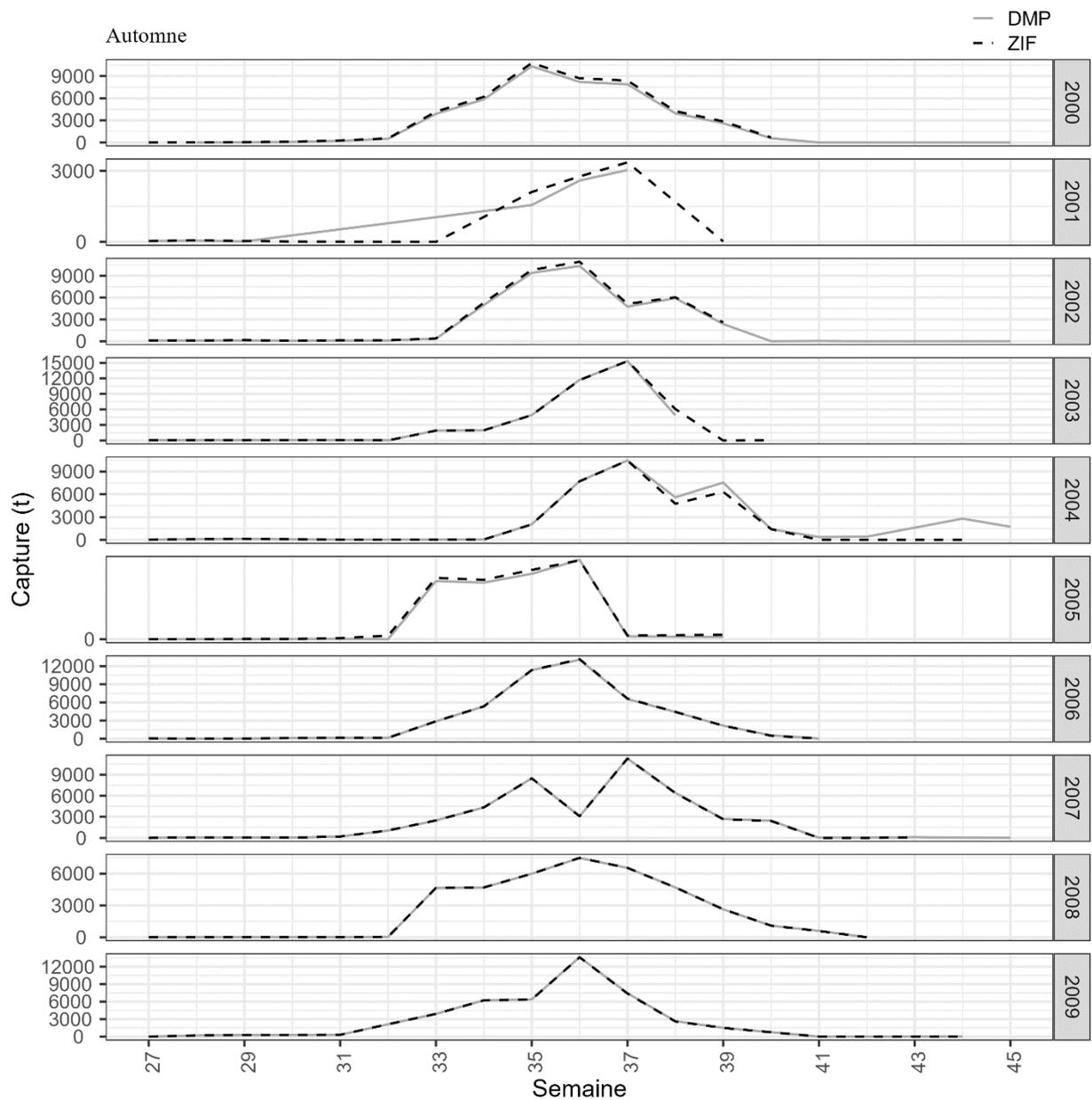


Figure 38. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche automnale entre 2000 et 2009. DMP en gris. ZIFF en noir hachuré.

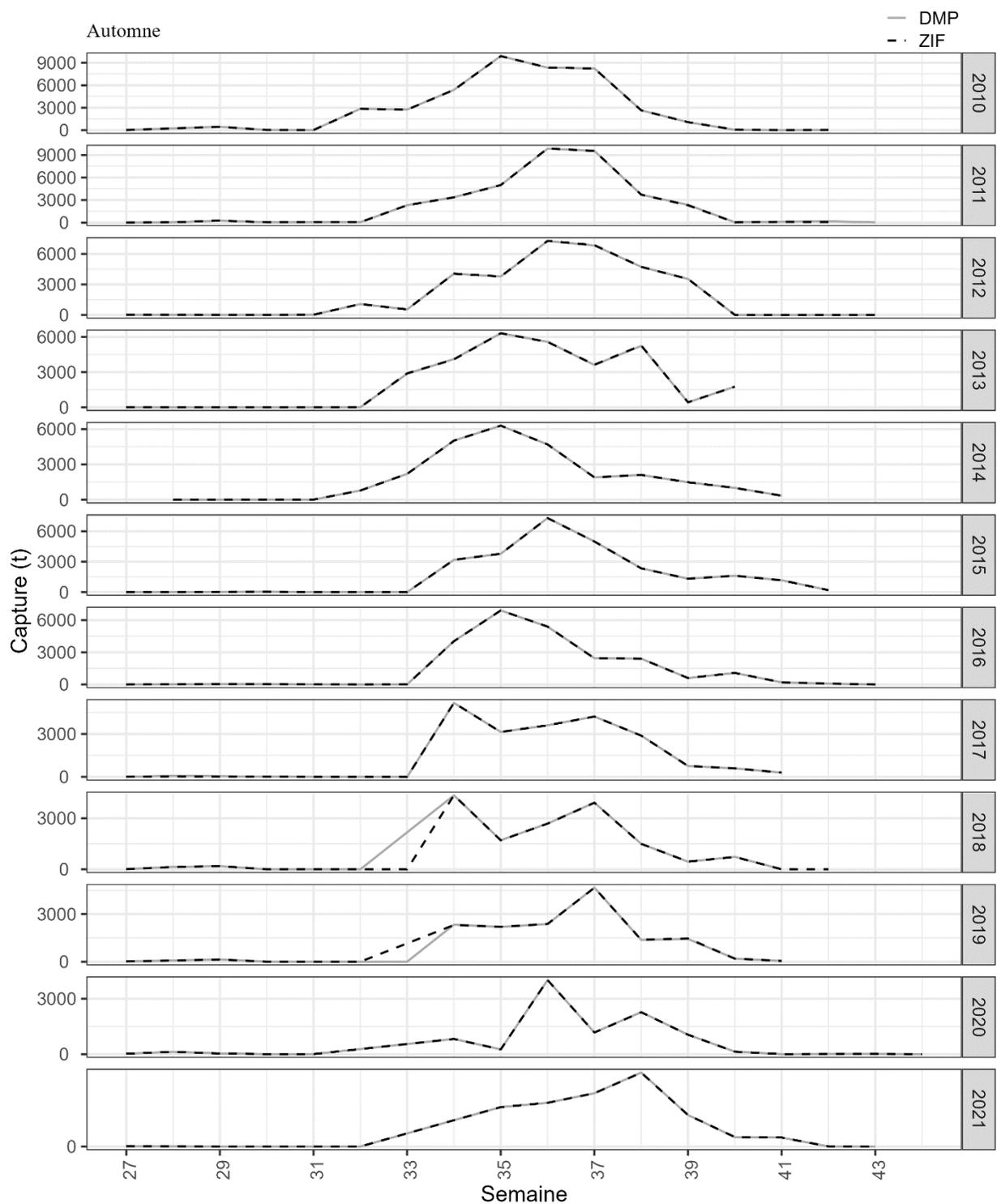


Figure 39. Capture en tonne par semaine pour les districts présents dans le DMP pour la pêche automnale entre 2010 et 2021. DMP en gris. ZIFF en noir hachuré.

La Figure 40 montre qu'il y a toujours plus de districts inclus dans les ZIFF que dans les données du DMP, mais aussi que certains districts sont représentés dans le DMP et absent des ZIFF pour la pêche automnale et printanière.

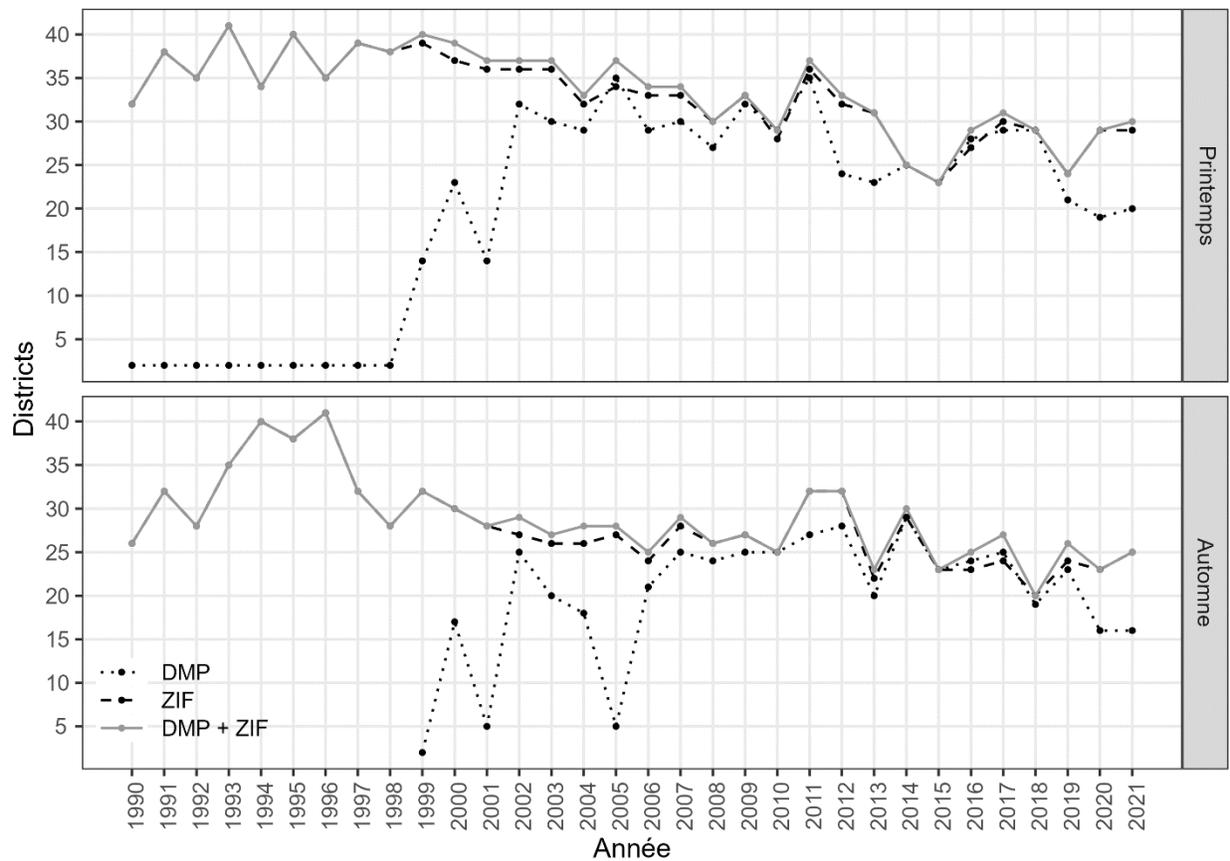


Figure 40 Comparaison des nombres de districts inclus pour les enregistrements du DMP (pointillé), des ZIFF (hachuré) et de la combinaison des 2 (trait plein) de 1990 à 2021 pour la pêche printanière et automnale.