



Pêches et Océans
Canada

Fisheries and Oceans
Canada

Sciences des écosystèmes
et des océans

Ecosystems and
Oceans Science

Secrétariat canadien des avis scientifiques (SCAS)

Document de recherche 2023/077

Région des Maritimes

Un cadre pour l'évaluation du crabe des neiges (*Chionoecetes opilio*) dans la région des Maritimes (divisions 4VWX de l'OPANO)

Jae S. Choi

Division de l'écologie des populations
Ministère des Pêches et des Océans
Institut océanographique de Bedford
C.P. 1006, Dartmouth (Nouvelle-Écosse) B2Y 4A2

Avant-propos

La présente série documente les fondements scientifiques des évaluations des ressources et des écosystèmes aquatiques du Canada. Elle traite des problèmes courants selon les échéanciers dictés. Les documents qu'elle contient ne doivent pas être considérés comme des énoncés définitifs sur les sujets traités, mais plutôt comme des rapports d'étape sur les études en cours.

Publié par :

Pêches et Océans Canada
Secrétariat canadien des avis scientifiques
200, rue Kent
Ottawa (Ontario) K1A 0E6

<http://www.dfo-mpo.gc.ca/csas-sccs/>
csas-sccs@dfo-mpo.gc.ca



© Sa Majesté le Roi du chef du Canada, représenté par le ministre du
ministère des Pêches et des Océans, 2023

ISSN 2292-4272

ISBN 978-0-660-67936-5 N° cat. Fs70-5/2023-077F-PDF

La présente publication doit être citée comme suit :

Choi, J.S. 2023. Un cadre pour l'évaluation du crabe des neiges (*Chionoecetes opilio*) dans la région des Maritimes (divisions 4VWX de l'OPANO). Secr. can. des avis sci. du MPO. Doc. de rech. 2023/077. v + 106 p.

Also available in English:

Choi, J.S. 2023. *A framework for the assessment of Snow Crab (Chionoecetes opilio) in Maritimes Region (NAFO Div 4VWX). DFO Can. Sci. Advis. Sec. Res. Doc. 2023/077. v + 101 p.*

TABLE DES MATIÈRES

RÉSUMÉ.....	v
INTRODUCTION	1
BIOLOGIE DU CRABE DES NEIGES.....	1
COLLECTE DES DONNÉES	3
ANALYSE DES DONNÉES.....	4
Agrégation spatio-temporelle	4
Synthèse	8
ESTIMATION DE L'INDICE D'ABONDANCE (AGRÉGATION DES DONNÉES)	8
MODÈLES SPÉCIAUX DE LISSAGE	9
MODÈLES STATIQUES	10
MODÈLES D'AUTOCORRÉLATION CONTINUE.....	10
Autocorrélation spatiale.....	10
Autocorrélation temporelle	12
Estimation spectrale de l'autocorrélation	13
Modèle spatial	15
Modèle spatial hiérarchique	16
Modèles spatio-temporels stationnaires.....	17
Modèles spatio-temporels complexes	18
MODÈLES HYBRIDES CARTÉSIENS-LAGRANGIENS	20
Modèles autorégressifs conditionnels spatiaux.....	21
Modèles spatio-temporels autorégressifs conditionnels.....	23
COMPARAISON DES RÉSULTATS	23
MODÈLES NON SPATIAUX ET NON TEMPORELS	24
MODÈLES D'AUTOCORRÉLATION CONTINUE.....	25
MODÈLES AUTORÉGRESSIFS CONDITIONNELS	29
Modèle Poisson-CAR.....	29
Variabilité des écosystèmes sur le plan des variables statiques	30
Variabilité des écosystèmes sur le plan des variables dynamiques	31
SYNTHÈSE.....	34
CONCLUSIONS.....	35
CODA.....	35
Modèles d'habitat (aussi appelés « modèles de répartition des espèces »).....	35
TABLEAUX	37
FIGURES	38
RÉFÉRENCES CITÉES	84
ANNEXE 1 : BIAIS D'ÉCHANTILLONNAGE	87
GRAPHIQUES DES RÉSIDUS	89

ANNEXE 2 : RÉSULTATS DU MODÈLE.....	98
A2.1. RÉSULTATS DU MODÈLE « FACTORIEL CROISÉ ». IL S'AGIT DU MLG UTILISANT L'INLA POUR ESTIMER L'ABONDANCE AVEC UNE HYPOTHÈSE DE RÉPARTITION DE POISSON, L'ESPACE ET LE TEMPS ÉTANT TRAITÉS COMME DES FACTEURS. IL N'Y A PAS D'AUTOCORRÉLATION.....	98
A2.2. RÉSULTATS DU MODÈLE GLOBAL STMV F() POUR LA TAILLE DES GRAINS DU SUBSTRAT DÉRIVÉS DES MODÈLES ADDITIFS GÉNÉRALISÉS.....	98
A2.3. RÉSULTATS DU MODÈLE « SIMPLE À EFFETS MIXTES ».....	99
A2.4. RÉSULTATS DU MODÈLE « STATIQUE À EFFETS MIXTES ».....	99
A2.5. RÉSULTATS DE L'ESTIMATION DE LA TEMPÉRATURE AU FOND – CAR POUR CHAQUE ANNÉE.....	100
A2.6. RÉSULTATS DE L'AXE 1 DE LA COMPOSITION DE L'ESPÈCE – CAR POUR CHAQUE ANNÉE.....	101
A2.7. RÉSULTATS DE L'AXE 2 DE LA COMPOSITION DE L'ESPÈCE – CAR POUR CHAQUE ANNÉE.....	102
A2.8. RÉSULTATS DU MODÈLE « DYNAMIQUE À EFFETS MIXTES ».....	103
A2.9. RÉSULTATS DU MODÈLE « SÉPARABLE ».....	104
A2.10. RÉSULTATS DU MODÈLE « SIMPLE NON SÉPARABLE ».....	104
A2.11. RÉSULTATS DU MODÈLE « ESPACE TEMPS NON SÉPARABLE ».....	105
A2.12. RÉSULTATS DU MODÈLE « HABITAT ESPACE TEMPS NON SÉPARABLE » POUR LE PETIT CRABE IMMATURE (LARGEUR DE CARAPACE < 50 MM).....	106

RÉSUMÉ

Le présent document est un examen de la modélisation spatio-temporelle dans le contexte de l'évaluation des stocks de crabes des neiges (*Chionoecetes opilio*) dans la région des Maritimes. Nous discutons de la justification de l'utilisation de la fonction d'autocorrélation et des modèles d'unités spatiales autorégressives conditionnelles. Nous démontrons l'utilité de ces derniers pour estimer l'abondance dans les programmes d'échantillonnage comportant des échantillonnages incomplets ou des défaillances sur le plan logistique. Ces considérations suggèrent l'importance de fonder les évaluations futures sur des modèles d'unités spatiales. Elles suggèrent également qu'une application systématique de ces méthodes dans toutes les espèces permettrait d'établir des comparaisons significatives de l'utilisation interspécifique de l'espace et des exigences ou sensibilités des écosystèmes. Elles sont aussi facilement adaptées aux scénarios pronostiques de changement climatique. Il s'agit d'étapes simples, mais nécessaires pour une gestion des pêches véritablement écosystémique.

INTRODUCTION

Le présent document vise à fournir un cadre pour l'évaluation des stocks de crabes des neiges (*Chionoecetes opilio*) dans la région des Maritimes du Canada atlantique (divisions statistiques 4VWX de l'OPANO; figure 1; ci-après, *évaluation du crabe des neiges*). La plupart des évaluations officielles des stocks comportent les éléments clés suivants :

- Collecte de données
 - Relevés pour le nombre, le poids, le genre et les variables environnementales
 - Assurance de la qualité et contrôle de la qualité (AQ/CQ)
 - Assimilation dans un entreposage à long terme
- Analyse des données
 - *Agrégation spatiotemporelle* à une zone fixe pour estimer les tendances temporelles à l'aide d'un plan expérimental (relevé)
 - *Synthèse* du rendement des pêches, des indices d'après le relevé, de la variabilité des écosystèmes et des contraintes biologiques pour fournir des projections et des analyses des risques

Plus précisément, l'étape de *l'analyse des données (agrégation spatiotemporelle)* ci-dessus a présenté un défi dans les évaluations précédentes du crabe des neiges. Dans la section qui suit, nous nous concentrerons sur cet élément en décrivant certaines de ces difficultés et en proposant une solution à la fois robuste et opérationnelle.

BIOLOGIE DU CRABE DES NEIGES

Le crabe des neiges a un cycle biologique complexe. Cette complexité entraîne une complexité spatiotemporelle dans ses répartitions d'abondance et est donc la principale cause des difficultés analytiques associées à *l'agrégation spatiotemporelle*. Le crabe des neiges est d'abord et avant tout une espèce *sténotherme* d'eau froide. Dans la région des Maritimes, le crabe des neiges pêché commercialement est observé à des profondeurs de 50 à 300 m et à des températures de -1 à 7 °C. On estime que le crabe des neiges évite les températures supérieures à 7 °C, car il a été démontré que les coûts métaboliques correspondent aux gains métaboliques près de cette température (Foyle *et al.* 1989). Cela signifie que leurs répartitions spatiales peuvent fluctuer de façon saisonnière et annuelle en raison de la nature complexe de la région des Maritimes sur le plan thermique. Les crabes plus petits et les femelles présentent également des différences dans les préférences thermiques. De plus, le crabe des neiges est fréquemment observé sur les fonds vaseux mous. On trouve également des crabes de petite taille et des crabes en période de mue sur des substrats plus complexes (rochers, cailloux), probablement parce que ceux-ci offrent plus d'abris (Sainte-Marie et Hazel 1992; Comeau *et al.* 1998). Bien sûr, de nombreux autres facteurs régissent leurs répartitions et leurs densités locales (voir ci-dessous). Le résultat final de ces interactions est que leur créneau environnemental variera selon le moment de l'année et l'endroit.

Les œufs de crabe des neiges sont couvés par leur mère pendant une période allant jusqu'à deux ans. Cela dépend aussi directement des températures ambiantes ainsi que de la disponibilité des aliments (qui dépend aussi de la température par la production primaire et secondaire) et de l'état de santé et de l'état de maturité de la mère (jusqu'à 27 mois chez les femelles primipares — premier événement reproducteur, et jusqu'à 24 mois chez les femelles multipares — deuxième ou troisième événement reproducteur; Sainte-Marie 1993). On a observé un développement plus rapide des œufs (de 12 à 18 mois) dans d'autres systèmes (Elnor et Beninger 1995, Webb *et al.* 2007). On estime que plus de 80 % des crabes des neiges

femelles sur la plate-forme néo-écossaise suivent un cycle annuel plutôt qu'un cycle bisannuel possiblement en raison de l'exposition à des températures plus chaudes (Kuhn et Choi 2011). Une femelle primipare ayant une largeur de carapace d'environ 57,4 mm produit entre 35 000 et 46 000 œufs qui sont expulsés entre février et avril (Sainte-Marie 1993). Les femelles multipares sont considérées comme plus fécondes, avec plus de 100 000 œufs produits par femelle. Les œufs éclosent d'avril à juin, lorsque les larves sont libérées. Ce stade larvaire pélagique dure de trois à cinq mois (stades zoé 1 et 2), et pendant cette période le crabe des neiges se nourrit de zooplancton. Les stades pélagiques semblent avoir la plus grande abondance en octobre et pourraient donc commencer à s'établir en janvier. Par la suite, ils s'installent plus près du fond au stade mégaloïde. On en sait très peu sur les taux de survie à ces premiers stades du cycle biologique, mais on estime que ces taux dépendent fortement de la température.

Une fois installé au fond (phase benthique), le crabe des neiges mue environ deux fois par année (Sainte-Marie *et al.* 1995, Comeau *et al.* 1998; figure 2), et les intervalles de fréquence et de taille de la mue dépendent également de la température et de la disponibilité des aliments. La première étape entre les mues (1^{er} stade) correspond à une largeur de carapace d'environ 3 mm. Après le 5^e stade (largeur de carapace de 15 mm), la fréquence de mue diminue aux mues printanières annuelles jusqu'à ce qu'elle atteigne une mue terminale à maturité. La croissance est allométrique et le poids augmente d'environ 250 % avec chaque mue. On a observé que la mue terminale se produisait entre le 9^e et le 13^e stade chez les mâles, et au 9^e et au 10^e stade chez les femelles. Juste avant la mue terminale, le crabe mâle peut sauter une mue une année et muer l'année suivante (Conan *et al.* 1992). Le crabe des neiges mâle atteint généralement la taille légale (\geq largeur de carapace de 95 mm) au 12^e stade, mais une fraction variable du crabe des neiges au 11^e stade a déjà atteint la taille réglementaire. Les crabes des neiges mâles au 12^e stade ont un âge d'environ 9 ans après la fixation des larves au fond et de 11 ans après l'extrusion des œufs. Par la suite, l'espérance de vie d'un mâle est d'environ cinq à six ans. Jusqu'à dix mois sont nécessaires pour que la carapace durcisse et jusqu'à un an pour que le rendement de la chair soit commercialement viable. Après le durcissement de la carapace, le mâle est capable de s'accoupler. Vers la fin de la durée de vie d'un crabe des neiges, la coquille se décalcifie et se ramollit, et pourrait présenter une forte croissance d'épibiontes.

Les femelles qui se reproduisent pour la première fois (femelles primipares) commencent généralement leur mue à maturité à une largeur moyenne de 60 mm de carapace et s'accouplent pendant que leur carapace est encore molle (début du printemps : avant la saison de pêche dans l'est de la N.-É. et pendant la saison de pêche dans la division 4X). Une deuxième période d'accouplement plus tard dans l'année (de mai à juin) a également été observée chez les femelles multipares (Hooper 1986). La formation de couples (étroite de la femelle par le mâle) peut avoir lieu jusqu'à trois semaines avant l'accouplement. Pendant l'accouplement, on a également observé chez le crabe des neiges des comportements complexes : le mâle aide la femelle primipare à muer, la protège des autres mâles et des prédateurs et va même jusqu'à la nourrir (indirectement). On a vu aussi, lors de l'émergence des larves, des mâles secouer les femelles pour aider à la dispersion des larves, c.-à-d., avant un accouplement multipare). Les femelles sont sélectives dans le choix de leur partenaire, comme c'est souvent le cas chez les espèces qui présentent un dimorphisme sexuel, et il arrive qu'elles trouvent la mort en résistant aux tentatives d'accouplement des mâles non élus (Watson 1972, Hooper 1986). Les mâles se disputent âprement la faveur des femelles, et se blessent souvent en le faisant (perdant certains de leurs appendices). Les grands mâles ayant de grandes pinces réussissent généralement mieux à s'accoupler et à protéger les femelles.

Le crabe des neiges a un cycle biologique complexe, est sensible à l'environnement et présente des variations ontogénétiques de cette sensibilité. C'est une espèce longévive qui est exposée à une diversité d'écosystèmes (biotiques et abiotiques) qui s'étendent sur une gamme d'échelles spatiotemporelles, et est capable de parcourir de grandes distances (en moyenne < 10 km, mais jusqu'à 100 km par année) pour la recherche de nourriture et la reproduction et des interactions interspécifiques complexes et très variées (figure 3). Cela rend les techniques d'agrégation spatiale statiques non viables.

COLLECTE DES DONNÉES

La pêche du crabe des neiges dans l'est du Canada a débuté en 1960, sous forme de captures accessoires par les dragueurs de poisson de fond près de Gaspé, au Québec. Dans la région des Maritimes, la pêche du crabe des neiges est active depuis le milieu des années 1970. Une évaluation scientifique du statut du crabe des neiges dans la région a débuté au milieu des années 1990 sur le bateau de pêche *Marco Michel* (Biron *et al.* 1997). Cette évaluation a été financée par l'industrie et élaborée par le Centre des pêches du Golfe (CPG, Moncton, Nouveau-Brunswick). Le mandat d'évaluation a été transféré à l'Institut océanographique de Bedford (IOB, Dartmouth, Nouvelle-Écosse) en 2004 (Choi *et al.* 2005b) et il continue d'être financé par l'industrie à ce jour. L'industrie considère qu'il s'agit d'un investissement visant la durabilité à long terme de la ressource. Avant 2004, des relevés étaient effectués d'avril à juillet; toutefois, cette période a été passée de septembre à décembre, principalement pour réduire le chevauchement avec la saison de pêche (avril à septembre, à l'exception de la zone CFA4X d'octobre à mars). De 2004 à 2013, des relevés ont été effectués sur le bateau de pêche *The Gentle Lady* jusqu'à son naufrage en décembre 2013. Les relevés de 2014 à aujourd'hui ont été effectués sur le bateau de pêche *Ms. Jessie*. Tout au long de ces périodes de transition, tous les efforts ont été déployés pour s'assurer que les caractéristiques des navires, l'équipage et le protocole d'échantillonnage étaient aussi uniformes que possible. Mais évidemment, un biais dépendant des navires est possible et probable.

Un chalut *Bigouden Nephrops*, un filet initialement conçu pour creuser dans des sédiments mous pour capturer des homards européens, a été utilisé pour échantillonner le crabe des neiges et d'autres espèces benthiques/démersales (Biron *et al.* 1997). Il possède une ralingue de 20 m, avec un maillage de 80 mm dans les ailes et de 60 mm dans le ventre et de 40 mm dans le cul de chalut. La configuration du filet varie selon les conditions locales. Elle a donc été enregistrée à l'aide de capteurs de surveillance du chalut sans fil, la profondeur et la température ont été enregistrées à l'aide de capteurs sans fil, et l'information sur la position a été enregistrée à l'aide de systèmes de géolocalisation. La durée réelle du contact avec le fond a été évaluée à partir des profils de surveillance du chalut et de température/profondeur, avec une cible nominale d'une durée de trait de 5 minutes. La zone de relevé du filet a été calculée à partir de la distance balayée et de la largeur du filet surveillée. La vitesse du navire a été maintenue à environ deux nœuds. La longueur de la fune était d'environ trois fois la profondeur. Tous les efforts ont été déployés pour rendre ce protocole uniforme, mais la structure du filet a évolué pour rendre le rendement plus précis et exact (tablier de cul de chalut, utilisation de flotteurs et de poids pour maintenir la structure du filet plus uniforme, ajustage des portes, etc.) et amélioration de la précision et de l'exactitude de l'électronique. Certains biais peuvent être introduits ici, mais l'amélioration de l'exactitude et de la précision a été jugée utile.

Tous les crabes ont été dénombrés et mesurés à l'aide d'étriers, l'état de la coquille a été déterminé, la dureté des pinces a été mesurée à l'aide d'un duromètre et pesée à l'aide d'échelles à compensation de mouvement. Cette dernière méthode a permis de mesurer directement la biomasse plutôt que de se fier aux relations allométriques entre les parties du corps pour estimer la biomasse (cette dernière méthode était celle utilisée avant 2004). L'état

de maturité des mâles a été déterminé à partir d'une combinaison de stadification biologique et d'analyse morphométrique. Bien que la maturité physiologique ne coïncide pas directement avec le début de la maturité morphométrique, cette dernière est plus facile à déterminer. Cela est déterminé par une augmentation disproportionnée de la hauteur des pinces par rapport à la largeur de la carapace chez les mâles. La maturité des femelles est évaluée à partir d'une inspection visuelle directe des œufs ou du développement des gonades. Lorsque l'état de maturité était ambigu, il a été déterminé morphométriquement, car la largeur de l'abdomen (mesurée par la largeur du cinquième segment abdominal) augmente rapidement par rapport à la largeur de la carapace au début de la maturité morphométrique, ce qui facilite la couvaison des œufs. D'autres prises accessoires ont été identifiées et l'abondance numérique et la biomasse ont été évaluées, avec des sous-échantillonnages au besoin. La morphométrie n'a pas été incluse pour les espèces non ciblées en raison de contraintes de temps et de personnel en mer.

Dans l'évaluation du crabe des neiges, une solution ouverte, à source ouverte et transparente a évolué pour intégrer l'information provenant des dossiers électroniques et papier dans des systèmes de données modernes. Les données des relevés sont entrées directement dans le système d'exploitation relationnelle de base de données des observateurs en mer, tenu au MPO, IOB (Institut océanographique de Bedford, Dartmouth, Nouvelle-Écosse) avec la double saisie des données et des limites de l'étendue de la base de données pour réduire les erreurs de saisie. D'autres processus de traitement ultérieur au moyen de processus d'AQ/CQ mieux adaptés sont encodés dans un progiciel [en langage R](#) extensible et ouvert, notamment l'identification des valeurs aberrantes et des données extrêmes aux fins de suivi et de correction manuels dans la base de données de stockage. Les données tirées des capteurs électroniques sont stockées localement et post-traitées aux fins d'AQ/CQ en R et préparées en vue de leur intégration au processus d'évaluation. Certaines de ces données sont partagées avec d'autres évaluations (p. ex. les températures). Le traitement des données, l'imputation et d'autres analyses sont effectués à des postes de travail locaux à l'aide de la bibliothèque R ci-dessus. Les méthodes de collecte des données sont réévaluées de façon continue et prêtes à évoluer, mais les changements apportés au protocole de relevé au fil du temps sont maintenant normalisés.

ANALYSE DES DONNÉES

Agrégation spatiotemporelle

Le principal défi en matière d'évaluation des stocks consiste à déterminer les ressources limitées et le temps requis pour échantillonner un domaine spatiotemporel très vaste, souvent avec des limites spatiales mal définies et des outils imparfaits susceptibles à l'usure, dans le but d'obtenir une représentation d'un ou de plusieurs processus (*latents*) non observables, mais réels qui permettent de déterminer l'état d'une espèce. Dans le relevé sur le crabe des neiges, il y a une forte densité de stations par rapport à la plupart des autres relevés dirigés, avec environ 400 stations visitées chaque année et qui prennent jusqu'à quatre mois. Le domaine spatial de l'évaluation du crabe des neiges est d'environ $109\,120\text{ km}^2$, ce qui signifie que chaque station représente en moyenne 273 km^2 ! Le problème est que chaque station n'échantillonne que $0,0039\text{ km}^2$, soit un ratio de 1:70 000. De même, sur le plan temporel, une station est échantillonnée en moyenne environ 5 minutes, mais elle est censée représenter une année complète (525 600 minutes), soit un ratio de 1:105 120. En termes d'unités espace-temps, c'est un ratio de $1:7\,36 \times 10^9$, soit 1 partie sur 10 milliards! Si l'environnement du fond marin était « bien mélangé » (chémostat homogène), il suffirait de quelques échantillons. Malheureusement, ce n'est pas le cas. Cet échantillonnage plutôt rare de l'océan nous oblige donc à faire des hypothèses, et l'estimation de toute caractéristique d'intérêt devient un

équilibre entre l'information obtenue par l'intensité de l'échantillonnage, c'est-à-dire des estimations précises (faible variabilité) et exactes (sans biais) *par rapport* aux coûts en ressources et en temps associés à l'obtention de l'information.

Dans les évaluations des pêches, le *plan expérimental* d'un relevé encode ces hypothèses et est donc utilisé *religieusement* comme base pour l'agrégation des échantillons dans l'espace. Des conseils sur l'état des stocks sont habituellement demandés sur une base annuelle pour certaines régions spatiales administratives arbitrairement déterminées (dans le présent document, *unités spatiales* ou *US*). Ces US sont habituellement associées à des précédents historiques ayant trait à l'accès et à l'allocation et/ou à des tentatives de rationalisation de l'échantillonnage à partir de certains renseignements supplémentaires. Les stratégies d'échantillonnage de ces US peuvent aller de l'échantillonnage entièrement aléatoire en l'absence de renseignements supplémentaires à une forme quelconque de plan aléatoire stratifié (voir aussi l'annexe 1, Biais d'échantillonnage). Dans ce dernier cas, les échantillons sont choisis au hasard parmi les US qui sont caractérisées, *a priori*, par des *facteurs informatifs*. Un exemple de poissons de fond est présenté à la figure 4, où il est clair que les profondeurs étaient la principale considération. L'analyse de (co)variance (AN(C)OVA) est une application courante de cette stratification, qui consiste à *bloquer* numériquement la variabilité de ces *facteurs informatifs* qui ne sont pas au centre de l'étude (pour que l'étude soit plus rentable). Plus la variabilité dans les US (par rapport à la variabilité entre les US) de la variable focale est faible, plus la stratification réussit à contrôler ces effets « nuisibles ».

Nous espérons qu'un échantillon au sein d'une US soit représentatif de l'US (*c.-à-d.*, « bien mélangé » à l'intérieur des US). Si ces facteurs de nuisance changeaient lentement (p. ex. changements de profondeur en raison des processus tectoniques) ou très rapidement (p. ex. niveaux d'éléments nutritifs biochimiques en raison de la reminéralisation bactérienne) par rapport à l'échelle temporelle des caractéristiques biologiques d'intérêt (variation de l'abondance en raison de la dynamique des populations), il pourrait s'agir d'une hypothèse et d'une approximation utiles. Le problème est évidemment que ces facteurs de nuisance changent à des échelles espace-temps qui sont semblables à la caractéristique d'intérêt, ce qui fait que les échantillons ne sont pas toujours représentatifs de la région qu'ils sont censés représenter. Cela peut entraîner un manque de précision (variabilité élevée) et un manque d'exactitude (biais). À mesure que le nombre de facteurs de nuisance augmente, le nombre d'US nécessaires pour bloquer adéquatement ces facteurs, statistiquement au sens de l'analyse de la variance classique, augmente à des niveaux non viables (sur le plan du temps et des ressources nécessaires). Cette augmentation est exponentielle en termes de covariables, car plus d'un échantillon est requis dans chaque bloc de covariables et, dans le cas de deux dimensions spatiales, le nombre d'unités augmente d'un facteur de deux par rapport à toute réduction linéaire d'échelle et, avec l'ajout du temps, d'un facteur supplémentaire ou plus selon qu'une discrimination saisonnière ou diurne est nécessaire.

Par exemple, lorsque les caractéristiques informatives sont dynamiques, leur pertinence dans les US statiques peut entraîner une non-concordance avec les facteurs d'importance présumés. Autrement dit, il peut y avoir des *repliements* spatiaux ou temporels (que l'on appelle parfois heuristiquement des problèmes de mise à l'échelle ascendante et descendante) en ce sens qu'un échantillon est prélevé à une température très différente (par exemple) ou à un moment de l'année que la température moyenne globale de l'unité spatiale qu'il est censé représenter. Dans le relevé sur le crabe des neiges, la période d'échantillonnage s'étend sur quatre mois au cours desquels les conditions environnementales à la surface passent des conditions estivales aux conditions hivernales. Ce blocage ou cette approche *factorielle* ne fait que bloquer l'influence d'une poignée de ces facteurs superflus (dans l'exemple, seulement la profondeur) et les ignore autrement comme facteurs de *nuisance*, car ils ne sont généralement même pas

mesurés. Dans le présent document, nous parlerons de cette approche d'application d'un plan expérimental *statique* comme perspective *cartésienne*. Cette *perspective cartésienne* domine les évaluations des stocks et les publications sur les pêches en général, et elle est troublante pour les raisons que nous verrons ci-dessous.

En réalité, ces facteurs de nuisance sont très informatifs et facilitent la compréhension des processus focaux d'intérêt. S'ils sont soigneusement mesurés et traités, ils peuvent appuyer des prévisions plus précises et exactes. Rappelons que dans le cas du crabe des neiges, l'échantillonnage représente 1 partie sur 10 milliards. Des renseignements supplémentaires aident à améliorer l'échelle de cet échantillonnage. En outre, lorsqu'ils agissent à des échelles spatiotemporelles semblables à celles des processus focaux, ces facteurs peuvent devenir très influents, de sorte que leur omission peut réduire la précision et l'exactitude. Par conséquent, dans l'évaluation du crabe des neiges, nous adoptons ces facteurs plutôt que d'essayer d'ignorer toute la variabilité des processus qui influencent le biote. Dans le présent document, nous l'appellerons la *variabilité des écosystèmes*. Parmi les exemples de *variabilité des écosystèmes* dans le contexte marin, mentionnons les interactions des organismes avec les variations des températures ambiantes, des niveaux de lumière, du pH, du potentiel redox, de la salinité, de la disponibilité des aliments et des nutriments, des abris, de l'abondance des prédateurs, de la prévalence des maladies, etc. La caractérisation de la *variabilité des écosystèmes* est en effet difficile et fastidieuse en raison de leur nature multiple et interdépendante. Toutefois, comme elle influence directement et indirectement le statut d'une population, elle ne peut et ne doit pas être ignorée. C'est particulièrement le cas si cette information existe déjà ou est moins coûteuse à échantillonner abondamment que la densité de population.

En effet, cette *variabilité des écosystèmes* induit quelque chose de très important : des autocorrélations spatiotemporelles complexes dans l'abondance de l'organisme d'intérêt. Comme l'indique la *Première loi de la géographie de Tobler : tout interagit avec tout, mais deux objets proches ont plus de chances de le faire que deux objets éloignés* (Tobler 1970). Si ce n'était pas le cas, alors tout ce que nous voyons serait complètement chaotique et discontinu, sans zones localisées d'homogénéité. Des arguments semblables peuvent être invoqués pour le temps. Ainsi, le plan expérimental doit porter une attention particulière à la *variabilité des écosystèmes* et à l'autocorrélation spatiotemporelle qu'il induit. Si la stratégie consiste à « bloquer » les facteurs de nuisance, chaque US doit également être indépendante l'une de l'autre, c'est-à-dire sans autocorrélation spatiale (et temporelle). Si le plan expérimental d'un relevé est inadéquat pour garantir une telle indépendance entre les US, un modèle spatiotemporel approprié peut être utilisé pour tenter de corriger ces biais et amorcer le processus 1) d'amélioration itérative du plan expérimental ou 2) d'amélioration de la collecte de données nécessaire pour corriger analytiquement les biais induits par la *variabilité des écosystèmes*. Dans le présent document, nous ferons référence à ce point de vue comme d'une perspective *lagrangienne*, qui perçoit les US comme étant plus fluides en définition que la perspective *cartésienne* statique identifiée précédemment.

Dans les régions complexes sur le plan hydrodynamique comme les divisions statistiques 4VWX de l'OPANO (région des Maritimes du Canada atlantique) où un certain nombre de courants océaniques convergent et où des changements climatiques rapides sont en cours et des changements écosystémiques rapides et importants ont eu lieu (Choi *et al.* 2005a), l'influence de la *variabilité des écosystèmes* ne peut être ignorée en toute sécurité. Les relevés de crabe des neiges, initialement au milieu des années 1990, étaient exploratoires et avaient lieu dans des régions côtières d'activités de pêche connues (figure 5). Les relevés se sont rapidement étendus à des échelles englobant l'ensemble du plateau lorsqu'il est devenu évident que leur répartition spatiale était beaucoup plus grande et plus hétérogène qu'on le croyait

auparavant. Cette expansion du relevé ne s'est stabilisée qu'au début des années 2000. Par conséquent, il existe un biais spatial qui empêche toute analyse simple ou tout mécanisme naïf d'agrégation. Un biais supplémentaire est également possible avec l'utilisation d'un nouveau bâtiment de pêche en 2004 et de nouveau en 2014. De plus, dans la zone la plus au sud de la répartition du crabe des neiges (zone de pêche du crabe 4X), la couverture des relevés au chalut a été sporadique dans le passé, mais s'est stabilisée depuis 2004 (à l'exception de 2014).

Historiquement, l'approche statistique utilisée pour agréger/intégrer les observations échantillonnées était le krigeage (Biron *et al.* 1997), une méthode d'interpolation des données qui s'inspire de l'information trouvée dans la façon dont la variabilité change selon la distance d'un lieu donné (covariance/autocorrélation spatiale, Cressie 1993). La conception optimale du relevé pour cette approche est généralement considérée comme une grille uniforme. À cette fin, les emplacements des relevés du crabe des neiges ont été choisis au hasard dans une grille spatiale de 10 10 minutes (Biron *et al.* 1997). La justification de la taille de la grille choisie n'est pas claire, car 10 minutes en longitude n'est pas la même distance que 10 minutes en latitude, et la distance de 10 minutes en longitude varie selon la latitude. Cela rend chaque grille d'échantillonnage inégale en surface et exerce donc une influence inégale sur l'intégration dans l'espace. L'échantillonnage dans chaque grille est/était aussi au mieux pseudo-aléatoire. En effet, de nombreux emplacements ne sont pas directement accessibles aux chaluts. Il est connu que le substrat de la région des Maritimes est accidenté et rocheux et comporte des tranchées profondes à de nombreux endroits, des caractéristiques qui ne permettent pas de passer les filets de pêche sans dommages ou de forcer les filets à faire partie du substrat. On a donc introduit un biais dans la sélection du type de fond, un biais qui favorise les emplacements accessibles au chalut, lesquels présentent aussi généralement des profondeurs uniformes et des substrats mous de gravier ou de boue qui coïncident avec les habitats préférés du crabe des neiges. Cela a donné lieu à un échantillonnage préférentiel d'emplacements plus profonds, de températures plus froides et d'associations d'espèces très différentes (figure 6).

Les emplacements des relevés sont regroupés de façon complexe en raison de l'historique des désignations des zones de pêche dans la gestion des pêches, du nombre de permis et de quotas, et de la nature financée par l'industrie du relevé avec des niveaux de contribution variables. Éventuellement, cela a donné lieu à des agrégations de stations de relevé dans les zones de pêche principales, à l'absence d'échantillonnage dans les zones non principales, ce qui a encore une fois introduit un biais dans les zones de l'habitat du crabe des neiges. En effet, l'échantillonnage d'emplacements connus comme éphémères ou peu susceptibles de comprendre du crabe des neiges a été considéré par beaucoup dans l'industrie comme un gaspillage de leur financement. Même à ce jour, nous ne connaissons toujours pas toutes les limites de leur répartition spatiale et de leurs déplacements en raison de contraintes opérationnelles et logistiques (les chaluts servant aux relevés ne peuvent pas raisonnablement dépasser 350 m de profondeur) et nous nous concentrons plutôt sur les emplacements accessibles à la pêche qui ont leurs propres contraintes opérationnelles, de profondeur et de distance.

À l'heure actuelle, l'échantillonnage est aussi exhaustif et intensif que possible, limité par les profondeurs accessibles au chalut (350 m), afin de permettre une détermination objective des limites spatiales de la population de crabe des neiges, de l'information qui doit être connue si des estimations fiables de la biomasse et de la structure de la population (p. ex. taille, sexe, maturité) doivent être établies. La répartition spatiale du crabe des neiges est assez dynamique et peut donc rapidement se déplacer vers des endroits où on ne les trouve pas « traditionnellement ». En outre, les modèles de répartition des crabes immatures, à carapace molle, très vieux et femelles ne correspondent pas entièrement à ceux des mâles de taille

réglementaire. Les premiers sont considérés comme moins compétitifs et plus susceptibles à la prédation (Hooper 1986), et sont généralement observés dans des environnements ou des substrats plus couverts (gravier, roches; Comeau *et al.* 1998). L'échantillonnage axé uniquement sur les régions où de grands mâles à carapace dure sont présents à haute fréquence empêcherait l'estimation fiable de l'abondance relative de ces autres segments importants de la population de crabes.

Synthèse

Dans la plupart des évaluations des stocks, les observations échantillonnées sont regroupées dans l'espace. Les indices temporels agrégés d'abondance qui en résultent sont utilisés tels quels et, s'il y a suffisamment d'information, ils sont soumis à une modélisation par séries chronologiques (espace d'état) qui *synthétise* l'information sur le rendement des pêches et les contraintes biologiques/structurelles sous forme de mortalité naturelle, de croissance et de profils de recrutement des stocks. Le résultat final est une évaluation du statut (habituellement axée uniquement sur la « biomasse du stock reproducteur », c.-à-d. les femelles reproductrices matures) par rapport à certains points de référence biologiques estimés à partir du modèle d'évaluation du stock.

À la suite d'un certain nombre d'accords internationaux, une position ministérielle a été formulée (MPO 2006) pour adopter une approche de précaution en matière d'évaluation et de gestion des pêches. Ici, la précaution a été utilisée dans un sens très particulier pour utiliser de tels *modèles de pêche synthétiques* pour définir les points de référence biologiques et les règles de contrôle des prises qui utilisaient ces points de référence. Dans l'évaluation du crabe des neiges, un modèle simple de dynamique de la biomasse est utilisé pour répondre à cette exigence et estimer des points de référence biologiques comme la capacité de charge et les taux intrinsèques d'augmentation. Les règles de contrôle des prises (figure 7) ont été encadrées dans ce contexte en 2010 (Choi et Zisseron 2011) et demeurent l'approche de gestion actuelle.

ESTIMATION DE L'INDICE D'ABONDANCE (AGRÉGATION DES DONNÉES)

L'estimation de l'état d'un système y_{st} est un objectif intermédiaire dans une évaluation des stocks. Lorsque la taille du domaine dans le temps et l'espace de y_{st} est grande, elle ne peut pas être observée *directement*. Les observations *indirectes* à un petit sous-ensemble d'emplacements et de périodes Y_{st} sont mesurées dans un espace $\{(s, t) \in D \in \mathbb{R}^d \times \mathbb{R}^1\}$ de coordination dans le domaine D de la dimensionnalité $d + 1$ avec des emplacements spatiaux $\{s = 1 \dots S\}$ et temporels $\{t = 1 \dots T\}$, et sont utilisées pour en déduire y_{st} . Ici, nous nous concentrons sur le cas simple des dimensions spatiales $d = 2$. Dans le cas des unités spatiales et temporelles agrégées, ces observations existent dans un ensemble d'unités spatiales non chevauchantes $u = 1 \dots U$ et d'unités temporelles $v = 1 \dots V$. Pour connecter Y_{st} à y_{st} , nous supposons que les observations Y_{st} sont dérivées d'une *fonction* de répartition stochastique spatiotemporelle qui génère (statistiquement) les observations (c.-à-d. un champ aléatoire spatiotemporel) :

$$y_{st} \xrightarrow{\text{processus spatiotemporel}} Y_{st}$$

et donc les observations Y_{st} sont utilisées pour déduire l'état réel y_{st} .

Les évaluations des stocks se concentrent presque toujours uniquement sur un *processus temporel* Y_t qui est généré par un processus spatiotemporel intégré sur le plan spatial $y_t = \int y_{st} ds$:

$$y_t \xrightarrow{\text{processus temporel}} Y_t$$

Cette intégration est encodée dans le *schéma expérimental* et est presque toujours un *processus purement spatial*. Cela signifie que les processus spatiaux sont traités comme étant emboîtés dans les processus temporels. Le problème est qu'un *schéma expérimental spatial* ne peut représenter un processus spatiotemporel que si la composante temporelle est stationnaire (au moins de premier et de deuxième ordre). Cela est généralement faux. (En effet, même l'hypothèse qu'un seul *processus spatial* soit valide dans l'ensemble du domaine spatial d'intérêt à un moment donné est en soi problématique, comme nous le verrons ci-dessous). Une application naïve d'un schéma spatial à un problème spatiotemporel peut donc poser problème; cette hypothèse doit être examinée.

Pour faciliter l'examen de cette hypothèse, nous commençons par le formalisme d'un modèle linéaire généralisé (Banerjee *et al.* 2004) :

$$Y \sim f(\mu, \varepsilon \sigma^2),$$

$$g(\mu) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \boldsymbol{\varepsilon},$$

avec les compensations constantes $\mathbf{O} = (o_1, \dots, o_S)$, le cas échéant; une matrice $S \times V$ de covariables $\mathbf{x} = (x_{sv}) \in \mathfrak{R}^{S \times V}$; les paramètres des covariables $V \boldsymbol{\beta}$ avec une distribution à priori normale multivariée avec matrice des variances moyenne μ et diagonale Σ ; et une erreur résiduelle $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_S)$. (Au niveau conceptuel, nous utilisons des indices à gauche pour désigner les types de variables et réservons les indices à droite pour l'indexation). Le $f(\cdot)$ indique une famille exponentielle (binomiale, normale, Poisson) et $E(Y) = \mu$ avec une fonction de lien réversible $g(\cdot) = f^{-1}(\cdot)$. Dans le cas de la distribution normale, $Y \sim \text{Normal}(\mu, \varepsilon \sigma^2)$ et $\mu = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \boldsymbol{\psi}$ avec écart type $\varepsilon \sigma$ associé à l'erreur ε résiduelle. Dans le cas de la distribution binomiale, $Y \sim \text{Binomial}(\eta, \theta)$ et $\ln(\theta/(1-\theta)) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \boldsymbol{\psi}$, où η est le vecteur du nombre d'essais et θ le vecteur des probabilités de réussite dans chaque essai. Dans le cas de la distribution Poisson, $Y \sim \text{Poisson}(\mu)$ et $\ln(\mu) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \boldsymbol{\varepsilon}$. Il convient de souligner qu'il existe une hypothèse audacieuse selon laquelle les *résidus sont indépendants et identiquement distribués*, c'est-à-dire : $\boldsymbol{\varepsilon} \stackrel{iid}{\sim} f(\cdot)$. Par conséquent, si une autocorrélation existe, par exemple lorsqu'il y a autocorrélation spatiale ou temporelle ou spatiotemporelle, l'estimation des paramètres du modèle peut devenir biaisée, y compris les estimations de l'abondance.

MODÈLES SPÉCIAUX DE LISSAGE

Des méthodes de visualisation rapide, comme l'interpolation linéaire ou du carré à distance inversée ou une simple tentative de lissage fondée sur la fonction spline pour répondre de manière ponctuelle à la *Première loi de la géographie de Tobler*. Toutefois, ces méthodes présentent un risque élevé de produire des tendances spatiales biaisées, car elles se concentrent uniquement sur les effets de l'ampleur (premier ordre) et/ou font de solides hypothèses *ad hoc* comme les tendances d'erreur résiduelle ou l'emplacement et l'ordre des points d'inflexion. Puisque la production d'estimations biaisées est un anathème dans l'évaluation des stocks, l'application naïve de ces méthodes n'est pas recommandée comme outil d'évaluation et ne devrait être utilisée que comme outil de visualisation rapide. Habituellement, en raison de la simplicité des méthodes, la variation temporelle est également traitée de façon indépendante (c.-à-d. totalement ignorée ou du moins l'autocorrélation temporelle est ignorée).

MODÈLES STATIQUES

Ignorer complètement l'autocorrélation spatiotemporelle est peut-être l'approche la plus fréquente. Cela peut sembler le plus simple et sans hypothèse. Mais il s'agit en fait de l'approche qui pose les hypothèses les plus audacieuses sur les données, elle suppose que la première loi de Tobler n'est pas applicable, de sorte que le schéma expérimental de l'étude est statique, toujours correct, sans biais et très précis. Cela signifie que la répartition du crabe des neiges à l'intérieur des unités spatiales (c.-à-d. une grille spatiale de 10 10 minutes) et temporelles (5 minutes) devrait être régulière, tandis que les répartitions entre les unités spatiales et temporelles adjacentes sont complètement aléatoires et chaotiques. Il s'agit toutefois d'hypothèses non viables, car elles ne sont pas étayées par des observations. Les unités spatiales et temporelles adjacentes ont tendance à présenter une certaine forme d'autocorrélation. Bien sûr, on peut toujours invoquer le théorème de la limite centrale, et une certaine mesure de la tendance centrale d'un échantillon aléatoire fournirait une estimation de la densité et donc de l'abondance globale même dans cette situation. Mais les hypothèses implicites ne sont en fait pas simples, même si la partie explicite du modèle peut l'être.

Dans le contexte d'un modèle de régression, cette *perspective* fondamentalement *cartésienne* est équivalente à un modèle d'interaction seulement entre deux facteurs contrôlés (unité spatiale et unité temporelle, encodées dans la matrice-modèle x^T) avec une certaine variabilité autour de chaque moyenne :

$$\begin{aligned} Y_{uv} &\sim N(\mu_{uv} \ \varepsilon \ \sigma_{uv}^2) \\ g(\mu_{uv}) &= \mathbf{x}^T \boldsymbol{\beta} + \varepsilon_{uv} \\ \varepsilon_{uv} &\sim N(0 \ \varepsilon \ \sigma_{uv}^2) \end{aligned}$$

C'est-à-dire que tous les échantillons à l'intérieur de chaque unité spatiale (u) et temporelle (v) sont censés être également représentatifs (contiennent la même quantité d'informations, « bien mélangées »). L'erreur résiduelle à l'intérieur de chaque unité espace-temps est indépendante et identiquement distribuée (iid; c'est-à-dire *interchangeable*). L'hypothèse habituelle est qu'ils suivent une distribution normale : $\varepsilon_{uv} \stackrel{iid}{\sim} N(0 \ \varepsilon \ \sigma_{uv}^2)$ avec une fonction de lien réversible $g(\cdot)$, dans ce cas, l'identité, et la variance $\varepsilon \ \sigma_{uv}^2$. La valeur doit également être positive puisque nous avons affaire à des organismes. Les relations entre les unités spatiales et temporelles sont complètement ignorées, chaque unité espace-temps étant censée fonctionner de manière *totale* *indépendante* de l'autre. On s'appuie strictement sur la capacité du schéma expérimental de l'étude à fournir une représentation *précise* et *exacte* de la réalité. Le modèle stratifié aléatoire est l'approche normale par défaut dans la plupart des évaluations des stocks. Une faiblesse importante de cette approche est que si une étude est incomplète pour une raison quelconque, alors il n'y a pas d'information pour cet emplacement à ce moment-là. Aucune inférence ne peut être faite, car il n'y a pas de mise en commun des informations entre les unités spatiales et les unités temporelles.

MODÈLES D'AUTOCORRÉLATION CONTINUE

Une perspective *lagrangienne* ne tient pas compte des limites entre les unités spatiales (u) et temporelles (v) et se concentre plutôt sur les relations spatiotemporelles et les autocorrélations entre toutes les données dans l'espace s et le temps t .

Autocorrélation spatiale

Une fonction d'autocorrélation spatiale ρ indique comment la proportion de la variance spatiale $C(h=0) = \omega \ \sigma^2$ diminue lorsque la distance augmente h . Il s'agit de la fonction de covariance $C(h)$ mise à l'échelle par la variance totale $C(0)$, c'est-à-dire, $\rho(h) = C(h)/C(0)$. La fonction de

covariance spatiale $C(h) = C(s, s'; \theta)$ exprime la tendance des observations voisines à plus interagir entre elles que celles qui sont plus éloignées (Première loi de la géographie de Tobler). Ici, la distance, $h = \|s - s'\|$, où $\|\cdot\|$ indique une norme spatiale qui, dans $d = 2$ dimensions spatiales, est simplement la distance euclidienne, $h = (\Delta \text{ordonnée}^2 + \Delta \text{abscisse}^2)^{1/2}$. Les formules couramment utilisées sont les suivantes :

$$\begin{aligned}
 C(h)_{\text{Sphérique}} &= \begin{cases} \omega \sigma^2 \left(1 - \frac{3}{2}h/\phi + \frac{1}{2}(h/\phi)^3\right); & 0 < h \leq \phi \\ 0; & h > \phi \end{cases} \\
 C(h)_{\text{Exponentielle}} &= \omega \sigma^2 e^{-h/\phi} \\
 C(h)_{\text{Gaussienne}} &= \omega \sigma^2 e^{-(h/\phi)^2} \\
 C(h)_{\text{Puissance exponentielle}} &= \omega \sigma^2 e^{-|h/\phi|^p} \\
 C(h)_{\text{Matérn}} &= \omega \sigma^2 \frac{1}{2^{\nu-1} \Gamma(\nu)} (\sqrt{2\nu}h/\phi)^\nu K_\nu(\sqrt{2\nu}h/\phi),
 \end{aligned}$$

À distance zéro, $C(0) = \text{Cov}(Y_s, Y_s) = \text{Var}(Y_s) = \varepsilon \sigma^2 + \omega \sigma^2$ (c'est-à-dire la variance spatiale dans son ensemble, également appelée seuil), où $\varepsilon \sigma^2$ est l'erreur non spatiale et non structurée (également dite « indépendante des effets spatiaux »), $\omega \sigma^2$ est l'erreur structurée à l'échelle spatiale (également appelée seuil partiel), et $\omega \theta = \{\phi, \nu, p, \dots\}$ sont des paramètres spécifiques à la fonction, notamment ϕ le paramètre de *plage*. $\Gamma(\cdot)$ est la fonction Gamma et $K_\nu(\cdot)$ est la fonction de Bessel de deuxième espèce avec un lissage ν . La fonction de covariance de Matérn est fréquemment utilisée dans la documentation plus récente, car la forme de cette fonction est plus souple et connue pour être liée à un processus aléatoire spatial gaussien (figure 8). L'approche des modèles spatiotemporels de la variabilité ([stmv](#)) (voir Section : Modèles spatiotemporels complexes) permet une estimation locale de ces paramètres d'autocorrélation et de l'échelle spatiale.

L'utilisation de cette approche permet une définition cohérente et non équivoque de l'échelle spatiale des observations (Choi et al. 2018). Il est impératif de connaître l'échelle spatiale de certains processus pour procéder à une évaluation empirique. Cette échelle spatiale peut être définie comme la distance h à laquelle l'autocorrélation tombe à un niveau au-delà duquel la relation est minimale. Ici, nous utilisons $\rho(x) = 0.1$ comme seuil, c'est-à-dire là où la variance spatiale chute à 10 % de la variance spatiale totale. Cette échelle spatiale est informative. Lorsque cette *échelle spatiale* est petite, cela signifie que les processus à courte portée dominent (par rapport à l'échelle de l'ensemble du domaine) et vice versa. Par exemple, lorsqu'une caractéristique spatiale (p. ex., la répartition de l'abondance dans l'espace) présente des *échelles spatiales* caractéristiques courtes (c.-à-d., une grande variabilité spatiale à de courtes distances), les approches d'échantillonnage doivent respecter cette caractéristique et fonctionner de la même façon à ces échelles plus courtes ou même plus petites si nous voulons être en mesure de résoudre les tendances et de décrire correctement le sujet d'intérêt. De même, si une caractéristique spatiale est de longue portée et que l'on souhaite résoudre correctement de tels modèles, le protocole d'échantillonnage doit être également de longue portée pour résoudre le modèle plus large. Un programme d'échantillonnage beaucoup plus petit que l'échelle spatiale caractéristique serait avantageux, mais les avantages accumulés par rapport au coût de l'échantillonnage diminueraient rapidement, dans la mesure où le temps, l'effort et les ressources nécessaires augmentent généralement plus rapidement que tout avantage (p. ex., dans le cas le plus simple, si l'on considère seulement naïvement l'erreur type comme une mesure de l'avantage, elle augmenterait asymptotiquement avec l'augmentation de l'effort avec une puissance de $-1/2$).

Les méthodes fondées sur la fonction d'autocorrélation supposent une stationnarité de premier et de deuxième ordre, c'est-à-dire que la moyenne et la variance sont stables et constantes dans l'ensemble du domaine spatial. Ils sont aussi assez lents à calculer.

Autocorrélation temporelle

Les systèmes écologiques évoluent également dans un cadre temporel. Ainsi, à l'instar des considérations spatiales ci-dessus, il existe également une certaine échelle temporelle caractéristique sur laquelle les processus internes à une zone d'intérêt et à une période d'intérêt se produisent. L'exemple canonique est le changement d'une certaine quantité d'une période discrète à une autre. Cette notion d'autocorrélation temporelle à temps discret est le paramètre de pente ρ d'un tracé d'une variable en fonction de lui-même avec un décalage d'une unité de temps :

$$v_{t+1} = \rho v_t + \eta_t,$$

avec $\eta_t \sim N(0, \tau \sigma^2)$ et un paramètre d'autocorrélation temporelle (linéaire) ρ . Il s'agit d'un processus AR(1), où le 1 indique un retard d'une unité. Notons que la valeur attendue est 0 pour η_t , c'est-à-dire qu'elle est *stationnaire au premier ordre*. Des modèles plus complexes avec des moyennes mobiles et des retards supplémentaires peuvent également être définis. Collectivement, ces modèles sont appelés modèles AR, ARMA et ARIMA. La difficulté de ces formules de séries chronologiques d'autocorrélation est l'exigence d'une série de données complète sans données manquantes, sous la forme type.

Dans une approche complètement identique au cas spatial, une fonction d'autocorrélation temporelle peut être utilisée pour décrire la forme du schéma d'autocorrélation temporelle (Choi *et al.* 2018). Plus précisément, nous définissons un *processus aléatoire* temporel, y_t , c'est-à-dire une certaine *fonction aléatoire* latente, non observable, mais réelle, qui génère des observations $y_t \rightarrow Y_t$, où Y_t sont des observations quelconques référencées temporellement à un certain moment t , mesurées dans un espace de coordonnées dont le domaine D a une dimension 1, de sorte que $\{t \in D \in \mathfrak{R}\}$ avec $\{l = 1 \dots L\}$ emplacements temporels. La façon dont la variabilité de Y_t change en fonction de la norme (distance), $h = \|t - t'\|$, est la fonction d'autocorrélation temporelle $\rho(h)$. Cette dernière peut prendre n'importe quelle forme, y compris les fonctions d'autocorrélation spatiale.

La fonction de covariance, par exemple, lorsqu'elle est exprimée sous la forme d'un modèle de décroissance exponentielle contrôlé par le paramètre de plage de temps ϕ_t est :

$$C(t, t'; \tau, \theta) \stackrel{\tau}{=} \sigma^2 e^{-|h|/\phi_t}.$$

À la différence de temps zéro, $C(0) = \text{Cov}(Y_t, Y_t) = \text{Var}(Y_t) = \varepsilon \sigma^2 + \tau \sigma^2$ (c.-à-d., variance temporelle globale), où $\varepsilon \sigma$ est l'erreur non temporelle et non structurée, $\tau \sigma$ est l'erreur temporelle structurée. La *fonction d'autocorrélation temporelle* est définie comme la fonction de covariance mise à l'échelle par la variance globale : $\rho_t(h) = C(h)/C(0)$. De manière heuristique, l'*échelle d'autocorrélation temporelle* est définie comme la différence de temps à laquelle l'autocorrélation temporelle diminue de manière asymptotique jusqu'à un certain niveau faible. Dans le présent document, nous utiliserons le même seuil que l'étendue spatiale pratique, $\rho_t(h) = 0,1$, et nous y ferons référence comme à l'échelle temporelle à laquelle l'autocorrélation temporelle approche 0,1 (soit 10 % de la variance temporelle totale). L'approche stmv (voir ci-dessous) permet d'estimer ces paramètres d'autocorrélation locale et l'échelle temporelle.

Estimation spectrale de l'autocorrélation

Les transformées de Fourier décomposent toute fonction dans un domaine continu (p. ex., temps, espace) en une somme infinie de sinus et de cosinus (Fourier 1822). Les fonctions sinus et cosinus sont interprétées comme des amplitudes et des déphasages associés à une gamme infinie de fréquences dans le domaine spectral. Des algorithmes efficaces du point de vue du calcul des transformations rapides de Fourier ont été initialement mis au point par Gauss en 1805 (Heideman et al. 1985), puis perfectionnés par Yates (1937), Danielson et Lanczos (1942) et entièrement généralisés par Cooley et Tukey (1965). Cela a permis de réaliser des opérations dans le domaine spectral qui sont des ordres de grandeur plus rapides que leurs équivalents dans les domaines temporel ou spatial continus. Par exemple, pour un problème comportant n données, la transformation rapide de Fourier a une complexité de calcul d'ordre $\mathcal{O}(n \log_2(n))$. En revanche, une opération d'importance dans le contexte de la modélisation spatiotemporelle est l'inversion d'une matrice des variances-covariances $\Sigma_{n \times n}$ qui a une complexité de calcul d'ordre $\mathcal{O}(n^3)$. Cela représente une amélioration par un facteur de $n^2/\log_2(n)$, ce qui, même pour un problème simple comportant $n = 10^3$ emplacements de données, peut représenter une amélioration de la vitesse de calcul de l'ordre de 10^2 . Entre parenthèses, la transformée de Fourier discrète a une complexité de calcul intermédiaire d'ordre $\mathcal{O}(n^2)$.

Au-delà de la complexité de calcul, il existe deux autres caractéristiques de la transformée de Fourier qui sont particulièrement importantes dans le contexte de la modélisation spatiotemporelle. Le premier est connu sous le nom de théorème de Wiener-Khinchin (-Einstein, -Kolmogorov) (Wiener 1930, Khintchine 1934, Robertson et George 2012), qui relie la fonction d'autocorrélation d'un processus aléatoire stationnaire à la densité spectrale de puissance (également appelée spectre de puissance) du processus. Autrement dit, une estimation rapide de l'autocorrélation (et de la corrélation avec retard) d'un processus peut être obtenue à partir du spectre. Le deuxième aspect important est le théorème de **convolution** : la combinaison de deux fonctions dans le domaine continu devient une simple multiplication dans le domaine spectral. La convolution d'une fonction d'autocorrélation avec une représentation spectrale du processus spatiotemporel d'intérêt équivaut à une interpolation de lissage des noyaux respectant l'autocorrélation temporelle, spatiale, spatiotemporelle. Ces deux aspects, respectifs, permettent des représentations rapides et sans biais de la fonction d'autocorrélation et une estimation ou prédiction rapide sans avoir à inverser et résoudre la matrice des variances-covariances $\Sigma_{n \times n}$.

Dans ce qui suit, nous nous concentrerons sur un problème unidimensionnel par souci de simplicité, tout en sachant que cela peut être simplement étendu à des dimensions supérieures, y compris les dimensions d'espace et d'espace-temps. Plus précisément, toutes les mesures le long de l'espace de coordination unidimensionnel $d = 1 \{(x) \in D \in \mathfrak{R}^d\}$, de domaine D , générées par le processus d'intérêt $g(x)$ peuvent être représentées dans le domaine des fréquences comme une série de coefficients trigonométriques complexes $G(k)$ et vice versa. Les transformées de Fourier directe et inverse sont respectivement :

$$G(k) = \mathcal{F}_x[g(x)](k) = \int_{-\infty}^{\infty} g(x) \cdot e^{-(2\pi ki)x} dx,$$
$$g(x) = \mathcal{F}_k^{-1}[G(k)](x) = \int_{-\infty}^{\infty} G(k) \cdot e^{(2\pi xi)k} dk.$$

Les éléments ci-dessus utilisent la formule d'Euler, $e^{2\pi\theta i} = \cos(2\pi\theta) + i \cdot \sin(2\pi\theta)$, pour représenter de manière condensée les amplitudes et les déphasages des fonctions sinus et

cosinus d'amplitudes $g(x)$ et $G(k)$, également appelées séries de Fourier. Le i représente par convention, des chiffres imaginaires.

Dans un contexte appliqué, la forme discrète de la transformée est particulièrement utile, car les mesures sont généralement discrètes à un certain niveau fondamental (comme un échantillonnage). Voici la transformée de Fourier discrète et son inverse :

$$G_k = \mathcal{F}_x[g_x](k) = \sum_{n=0}^{N-1} g_x \cdot e^{-(2\pi ki)(x/N)},$$

$$g_x = \mathcal{F}_k^{-1}[G_k](x) = \frac{1}{N} \sum_{k=0}^{N-1} G_k \cdot e^{(2\pi xi)(k/N)}.$$

Les valeurs $g_x = \{g_0 \ g_1 \ \dots \ g_{N-1}\}$ sont des vecteurs de valeurs dans le temps ou l'espace des données et les valeurs $G_k = \{G_0 \ G_1 \ \dots \ G_{N-1}\}$ sont des bandes de fréquences discrètes. Pour chaque bande de fréquences k , l'amplitude associée est :

$$|G_k|/N = \sqrt{\text{Re}(G_k)^2 + \text{Im}(G_k)^2}/N$$

et la phase est l'angle $\arg(G_k)$ entre la ligne réelle et la ligne imaginaire :

$$\arg(G_k) = \text{atan2}(\text{Im}(G_k) \ \text{Re}(G_k)) = -i \cdot \ln\left(\frac{G_k}{|G_k|}\right),$$

L'utilité de la représentation spectrale est que la fonction d'autocorrélation $\rho(x)$ d'une certaine fonction **stationnaire** $g(x)$ est équivalente à la transformée de Fourier inverse de la répartition spectrale de puissance.

$$\rho(x) = \mathcal{F}_k^{-1}[|G_k|^2](x)$$

où, $|G_k|^2$ est la densité spectrale de puissance au carré du module dérivé de :

$$\mathcal{F}[g(x) * g(-x)](k) = G_k \cdot G_k^* = |G_k|^2$$

et l'astérisque indique un nombre complexe conjugué.

Cette relation est le résultat du théorème de Wiener-Khinchin (Robertson et George 2012). La fonction d'autocorrélation est évidemment directement liée à la fonction de covariance utilisée dans les méthodes d'interpolation temporelle, spatiale et spatiotemporelle. L'utilité n'est toutefois pas simplement une question de complexité de calcul considérablement réduite de $\mathcal{O}(n^3)$ à $\mathcal{O}(n \log_2(n))$. Les méthodes d'estimation par semi-variogramme ou autocorrélation font généralement un usage répété des données pour l'estimation de la variance dans divers groupes de distance, ce qui entraîne une fausse autocorrélation et, par conséquent, des estimations biaisées des paramètres. La fonction d'autocorrélation, étant dérivée d'une décomposition spectrale, est, par définition, indépendante et donc les estimations des paramètres sont sans biais!

En outre, ces méthodes fondées sur la transformation rapide de Fourier nécessitent un traitement supplémentaire parce que les données manquantes sont courantes dans la plupart des applications empiriques. Ceci peut être traité par un schéma de moyenne pondérée localement (poids du noyau) qui, dans le domaine spectral, est connu sous le nom de lissage de convolution à noyau de Nadaraya-Watson (voir, par exemple, la documentation dans la librairie R, fields, Nychka et al. 2017). Il convient de rappeler que ces approches basées sur la transformation rapide de Fourier sont soumises aux mêmes difficultés que les approches basées sur la covariance, dans la mesure où la stationnarité (premier et deuxième ordre) est supposée ou requise (théorème de Wiener-Khinchin). L'approche stmv (voir la section : Les

modèles spatiotemporels complexes, ci-dessous) utilise ces méthodes fondées sur la transformation rapide de Fourier pour faciliter l'estimation des paramètres d'autocorrélation et des échelles spatiales et temporelles.

Modèle spatial

Le crabe des neiges présente des agrégations groupées dans l'espace et dans le temps. Cela signifie que l'erreur résiduelle à l'intérieur de chaque unité temporelle ne sera pas iid, a priori. En outre, les unités spatiales les plus proches ont tendance à être plus similaires que celles qui sont plus éloignées, et les échantillons d'une unité de temps donnée présentent également une corrélation. Les premières tentatives d'estimation de l'abondance du crabe des neiges ont été conçues en mettant l'accent sur la composante spatiale de l'erreur résiduelle. En ce qui concerne la formulation du modèle, il s'agit d'une extension progressive du *modèle* $\mathbf{Y}_t = (Y_{1t} \dots Y_{St})^T$ qui ajoute une erreur spatiale (ω_t) au modèle de base et traite ainsi chaque tranche de temps t , *indépendamment*. L'indice t pourrait être abandonné par souci de concision, mais nous le laissons explicitement pour nous rappeler qu'il s'agit d'un processus spatiotemporel où le temps est présumé indépendant :

$$Y_t \sim \text{MVN}(\mu_t, \Sigma_t)$$

$$g(\mu_t) = \mathbf{x}_t^T \boldsymbol{\beta}_t + \omega_t + \varepsilon_t$$

Ici, la composante non spatiale demeure la composante indépendante et identiquement distribuée, mais elle s'applique maintenant au domaine complet et non à une seule unité de domaine, tandis que la composante spatiale comporte une erreur qui change en fonction de la distance (la fonction d'autocorrélation) :

$$\varepsilon_t \sim N(0, \varepsilon \sigma_t^2)$$

$$\omega_t \sim \text{GP}(\mathbf{0}, C(s_t, s_t'; \omega, \theta_t))$$

La composante non spatiale de chaque année ε_t représente la variabilité des processus de mesure ou de microéchelle. Dans sa forme la plus simple (historique), on suppose généralement qu'elle a une distribution normale avec une moyenne de zéro et un écart type $\varepsilon \sigma$ (également appelée erreur *indépendante des effets spatiaux*). On suppose que l'erreur spatiale suit un **processus gaussien** avec une moyenne de zéro et une fonction de covariance spatiale $C(s_t, s_t'; \omega, \theta_t)$ qui décrit la forme de la variance du processus en fonction de la distance entre les données, contrôlée par les paramètres ω, θ_t et l'écart type spatialement structuré $\omega \sigma$ (voir ci-dessous). La matrice des variances-covariances est la somme de ces deux variances : $\Sigma_t = [C(s_{it}, s_{jt}; \omega, \theta_t)]_{i,j=1}^K + \varepsilon \sigma_t^2 I_S$, avec I_S une matrice identité de taille S .

Ce *modèle spatial* (aussi appelé *krigeage*) a été utilisé dans les premières évaluations du crabe des neiges. Cependant, cela a nécessité le calcul répété de la fonction d'autocorrélation pour chaque année (en fait, la matrice des variances-covariances, voir ci-dessous). Il convient de souligner que la forme de la fonction d'autocorrélation n'était pas constante. Elle variait en raison des variations de l'abondance spatiotemporelle et des distributions. Ce qui est particulièrement *alarmant*, c'est qu'il n'a pas été possible d'obtenir une estimation fiable d'une fonction d'autocorrélation lorsque les densités et les distributions spatiales se sont contractées (p. ex. en 2005), une situation où l'on a en fait besoin d'une estimation plus précise et exacte, plutôt que moins. En effet, il n'était pas évident de savoir si l'on devait modéliser la fonction d'autocorrélation spatiale elle-même en tant que caractéristique autocorrélée dans le temps, si l'on devait intégrer l'autocorrélation temporelle au modèle ou si l'on devait simplement utiliser par défaut une fonction d'autocorrélation pondérée dans le temps (toutes ces tentatives ont été réalisées).

Modèle spatial hiérarchique

L'ajout d'autocorrélation spatiale améliore notre compréhension de l'influence d'une station sur son environnement immédiat et de la façon dont elle se dégrade avec la distance en fonction des caractéristiques de variance. De cette façon, il s'agit d'une amélioration progressive par rapport à un modèle non spatial. Toutefois, l'hypothèse de l'ancien modèle est que l'erreur non structurée est de $\varepsilon \stackrel{iid}{\sim} N(0, \sigma_\varepsilon^2)$. Malheureusement, la répartition des données sur le crabe des neiges suggère que l'erreur est non gaussienne; elle est peut-être plus proche d'une distribution logarithmique. Cela s'explique en grande partie par les regroupements spatiaux découlant de facteurs comportementaux tels que les groupes de reproduction, l'évitement des prédateurs, la recherche d'aliments épars et les contraintes d'habitat telles que les limites de température et la disponibilité du substrat, ce que nous appelons naïvement la *variabilité de l'écosystème*. Le résultat de ces regroupements spatiaux est que la répartition *numérique* de l'abondance échantillonnée sous forme de grille ou dans tout schéma expérimental naïf sera également *non gaussien*. La plupart des échantillons manqueront les regroupements, ce qui fait que de nombreux échantillons sont des emplacements à faible densité et un nombre plus rare de stations avec regroupements à haute densité.

Le modèle de distribution des données est important, car le calcul de la moyenne arithmétique d'une distribution logarithmique n'est pas la meilleure estimation de la tendance centrale, car en raison de la forme de la distribution, il la surestimerait. Toutefois, la plupart des formes historiques de krigeage fonctionnent généralement selon l'hypothèse que la distribution des erreurs des données est normale (gaussienne). En d'autres termes, la nature groupée de l'abondance du crabe des neiges donnera lieu à des estimations spatiales de l'abondance qui sont généralement plus élevées que l'abondance réelle probable, simplement en raison des données et de la distribution spatiale. Ce ne serait pas une approche de précaution.

Il est donc essentiel de modifier notre hypothèse de distribution des données. Cela est facile à mettre en œuvre en modifiant l'hypothèse de distribution du *modèle spatial (ci-dessus)* :

$$\begin{aligned} Y_t &\sim \text{MVN}(\mu_t, \Sigma_t) \\ \log(\mu_t) &= \mathbf{x}_t^T \boldsymbol{\beta}_t + \omega_t + \varepsilon_t \\ \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ \omega_t &\sim \text{GP}(\mathbf{0}, C(s_t, s_t'; \omega, \theta_t)) \end{aligned}$$

pour la biomasse à valeur positive. C'est intéressant, car les résultats sont moins sujets aux biais induits par la distribution. Toutefois, comme il est mentionné ci-dessus, cela ajoute une autre forme de biais au modèle en ce sens que toutes les valeurs nulles sont ignorées (éliminées) puisque $\log(0)$ n'est pas défini. Une façon de contourner cette situation est d'avoir un modèle binomial distinct pour décrire les valeurs nulles et non nulles.

$$\begin{aligned} H_t &\sim \text{Binomial}(\eta_t, \theta_t) \\ \ln(\theta_t/(1 - \theta_t)) &= \mathbf{H}\mathbf{x}_t^T \mathbf{H}\boldsymbol{\beta}_t^H + \omega_t^H + \varepsilon_t^H \\ \varepsilon_t^H &\sim N(0, \sigma_\varepsilon^2) \\ \omega_t^H &\sim \text{GP}(\mathbf{0}, C(s_t, s_t'; \omega, \theta_t)) \end{aligned}$$

où l'exposant gauche H indique que le paramètre est obtenu à partir du modèle binomial (habitat), η_t est le nombre d'essais et θ_t les probabilités de réussite à chaque essai au cours d'une année donnée t et H_t est la proportion d'échantillons avec présence de crabe des neiges dans un endroit donné (probabilité d'habitat) et une année donnée. Ce dernier modèle devient un excellent modèle de probabilité d'habitat (également appelé modèle de distribution des espèces dans la documentation actuelle) qui peut fournir à lui seul de nombreuses informations

sur l'espace d'habitat du crabe des neiges. Lorsqu'il est combiné avec le modèle d'abondance, il devient un modèle hiérarchique gonflé à zéro. Cependant, cela entraîne des calculs supplémentaires et un doublement du nombre de paramètres, ainsi que la nécessité d'effectuer des simulations a posteriori pour propager correctement les erreurs et l'absence de moyen simple d'évaluer la performance du modèle dans son ensemble. [EN OUTRE : cette approche hiérarchique est également connue sous le nom de « modèles Delta » ou « modèles avec barrières »].

Modèles spatiotemporels stationnaires

Jusqu'à présent, les modèles n'ont utilisé qu'une représentation de l'autocorrélation spatiale. La plupart des caractéristiques biologiques comme l'abondance sont toutefois variables non seulement dans l'espace, mais aussi dans le temps. Tout comme l'espace peut être modélisé comme un processus autocorrélé, le temps aussi. Toutefois, en raison de la brièveté de la plupart des séries chronologiques biologiques, une fonction d'autocorrélation temporelle complète et fiable ne peut normalement être définie ou estimée. Au contraire, une extraction opérationnelle de quelques composantes, à savoir une composante annuelle et éventuellement une composante saisonnière, peut fournir certaines informations et contraintes sur les erreurs temporelles résiduelles. Le modèle le plus simple utilise une seule fonction d'autocorrélation spatiale pour l'ensemble du domaine et un seul paramètre ou une seule fonction d'autocorrélation temporelle (séparable, additive) (si les données de séries chronologiques sont suffisantes).

Les modèles spatiotemporels peuvent être considérés comme une simple extension du modèle de régression spatiale. Le modèle de régression espace-temps peut alors être défini comme suit :

$$\begin{aligned} Y &\sim \text{MVN}(\mu, \Sigma) \\ \log(\mu) &= \mathbf{x}^T \boldsymbol{\beta} + \omega + \varepsilon \\ \varepsilon &\sim \text{N}(0, \sigma^2) \\ \omega &\sim \text{GP}(\mathbf{0}, C(s, s'; \omega\theta)) \end{aligned}$$

où le processus d'erreur est décomposé en une composante structurée spatiotemporelle ω et une composante non structurée ε . La complexité réside dans la façon dont ces paramètres *spatiotemporels* sont paramétrés ($\omega, \varepsilon, \beta, \theta$, etc.). Par exemple, l'*erreur structurée dans l'espace* peut être paramétrée comme suit :

- ω – entièrement fixé
- ω_{-t} – effets spatiaux imbriqués dans le temps (autocorrélation spatiale à chaque tranche de temps, pas d'autocorrélation temporelle);
- ω_{s-} – effets temporels emboîtés dans les sites (autocorrélation temporelle à chaque site, pas d'autocorrélation spatiale);
- $\omega_{s-} + \omega_{-t}$ – *séparables* (les autocorrélations spatiales et temporelles sont indépendantes et additives ou multiplicatives si sur une échelle logarithmique) avec $\omega_{-t} \sim \text{GP}(\mathbf{0}, C(\mathbf{s}_t, \mathbf{s}_t'; \omega\theta_t))$ et $\omega_{s-} \sim \text{GP}(\mathbf{0}, C(t_s, t_s'; \tau\theta_s))$;
- ω_{st} – non séparable (la structure du temps et de l'espace évolue de manière non simple).

Chacune des diverses erreurs et chacun des divers paramètres peuvent avoir des complexités similaires, ce qui mène à une expressivité plutôt extraordinaire et à des exigences de calcul tout aussi extraordinaires. Bien qu'elle soit cohérente et élégante sur le plan conceptuel, l'évaluation

des vraisemblances dans ces modèles exige le calcul répété de l'inverse de la matrice $\Sigma_{n \times n}$ de covariance de taille n , une opération qui s'adapte aux opérations $\mathcal{O}(n^3)$. Cela a été un obstacle à l'élaboration ultérieure de ces méthodes fondées sur la covariance dans des problèmes à grande échelle en matière d'espace et d'espace-temps. Des approximations ont été proposées pour surmonter cette limite de calcul : modéliser le processus spatial ω au moyen d'un processus dimensionnel inférieur par l'intermédiaire de convolutions à noyaux, déplacer des moyennes, fonctions splines/de base de degré faible et processus prédictifs (projection d'un processus spatial sur un sous-ensemble plus petit; Sølna et Switzer 1996, Wikle et Cressie 1999, Huang et al. 2004, Xu et al. 2005, Banerjee et al. 2004); estimer approximativement le processus spatial comme un champ aléatoire de Markov avec les approximations de Laplace et d'EDPS (Lindgren et Rue 2015) et estimer approximativement la vraisemblance du processus spatiotemporel d'EDPS à l'aide d'un processus du domaine spectral (Sigrist et al. 2012).

Les modèles séparables sont presque toujours utilisés pour des raisons de vitesse de calcul, car ils traitent l'espace et le temps de façon indépendante, ce qui réduit grossièrement les problèmes des opérations $\mathcal{O}((ST)^3)$ à $\mathcal{O}(S^3) + \mathcal{O}(T^3)$; où S le nombre de positions spatiales et T le nombre de tranches de temps. En réalité, cependant, ces modèles séparables sont généralement inappropriés à moins que la zone d'étude ne soit homogène et vraiment constante de premier et de second ordre (c.-à-d. moyenne constante, variance dans le temps et l'espace), un fait qui est rarement vrai dans la plupart des systèmes écologiques. C'est notamment le cas en biologie où l'agrégation et le comportement dépendent fortement du contexte et des groupes (lieu et temps) (non linéaire/non additive). Il y a des discontinuités dans l'espace et le temps en raison du paysage, de l'effondrement d'espèces, des invasions, etc.

Modèles spatiotemporels complexes

Toutes ces approches, y compris la plupart des modèles spatiotemporels opérationnels, ont en commun une hypothèse de base, à savoir que la moyenne et la variance sont stables dans tout le domaine (stationnarité de premier et de second ordre) et qu'une seule fonction d'autocorrélation dans l'espace et une seule fonction d'autocorrélation dans le temps suffisent à décrire le modèle spatiotemporel complet. L'exactitude de ces hypothèses particulièrement importantes est rarement vérifiée dans la documentation (je ne les ai jamais rencontrées). Lorsqu'on les examine, il y a peu d'éléments qui permettent de penser qu'il s'agit d'une hypothèse appropriée.

Une façon de traiter cette non-stationnarité consiste à induire l'autocorrélation par l'entremise de covariables. En élargissant le modèle pour inclure des covariables variables dans le temps et statiques qui peuvent expliquer certaines de ces discontinuités, les distributions d'erreurs peuvent devenir plus stationnaires. C'est-à-dire élargir le contenu de $\mathbf{x}^T \boldsymbol{\beta}$ pour inclure les processus qui représentent la *variabilité de l'écosystème*. L'éventail des possibilités est bien sûr très large et comme chaque composante de la *variabilité de l'écosystème* nécessite généralement sa propre modélisation et évaluation, l'exercice peut très vite devenir onéreux.

Une autre solution consiste à identifier un sous-domaine local où ils sont effectivement stationnaires dans l'espace et éventuellement stationnaires dans le temps également. Cela réduit le problème à de petits sous-domaines gérables où les hypothèses de stationnarité locale sont valides et où la modélisation des processus spatiotemporels et des paramètres associés devient réalisable du point de vue informatique. Cette approche est mise en œuvre dans la librairie [stmv](https://github.com/jae0/stmv) (Choi et al. 2018; <https://github.com/jae0/stmv>). Il y a donc une certaine similitude conceptuelle entre cette approche et la « régression pondérée géographiquement » (p. ex. Fotheringham et al. 2002), car chaque sous-domaine peut avoir ses propres paramètres de modèle β_{st} . Cependant, la régression pondérée géographiquement ne permet de faire varier

que les paramètres du modèle β_{st} , or, ici, tant les paramètres du modèle β_{st} que les erreurs spatiotemporelles ω_{st} et ε_{st} varient également dans l'espace.

Pour être plus précis, nous définissons les nœuds statistiques $\{N_{m=(1 \dots M)} | m \in \mathfrak{R}^d\}$ dans un treillis spatial qui sont les centres des unités spatiales (figure 9). La norme (distance) des données de chaque nœud est $h_m = ||s_m - s_Y||$. Un sous-domaine local d'un nœud donné m est $\{S_{m=(1 \dots M)} \in D | h_m < h_u\}$ ou plus brièvement comme $\{S_m\}$ qui représente tous les emplacements situés à une certaine distance du nœud statistique $\{h_u | \rho(h_u)_{\text{Matérn}} > 0.1\}$; c'est-à-dire la distance à laquelle l'autocorrélation spatiale locale tombe à une valeur négligeable (prise arbitrairement comme $\rho > 0.1$) et les valeurs de paramètres associées sont prises en charge. Les données spatiotemporelles présentes dans ce sous-domaine m sont $\{Y_{st} | (s, t) \in D | h_m < h_u\}$ et sont ici abrégées sous la forme de ${}^m Y_{st}$.

L'utilisation de toutes les composantes du modèle de régression est prohibitive sur le plan du calcul. Même avec des paramétrages très simplistes de la structure spatiale et temporelle par le modèle additif généralisé (MAG) ou le modèle additif mixte généralisé (MAMG), les solutions prennent plusieurs semaines sur des machines rapides (CPU de 5 GHz, 64 Go de RAM en 2016), selon la quantité de données, la résolution et la complexité du modèle, bien sûr. Comme compromis entre la complexité du modèle et la vitesse de calcul, l'approche stmv utilise un modèle de covariable global $\mathbf{x}^T \boldsymbol{\beta}_{st} + \varphi_{st}$ peut être utilisé pour paramétrer les effets des covariables en utilisant un modèle linéaire, linéaire généralisé ou additif généralisé. Des fonctions splines de base pénalisées de faible ordre des prédicteurs de covariables de trois nœuds ou moins sont biologiquement plausibles, car on peut souvent s'attendre à une modalité :

$$\begin{aligned} Y &\sim N(\mu_{st}, \varphi, \sigma^2) \\ g(\mu_{st}) &= \mathbf{x}^T \boldsymbol{\beta}_{st} + \varphi_{st} \\ \varphi_{st} &\sim N(0, \varphi, \sigma^2), \end{aligned}$$

L'erreur spatiotemporelle résiduelle qui en résulte φ_{st} peut être décomposée en d'autres composantes dans une approche similaire à celle du krigeage par régression et du krigeage universel avec dérive externe (Hengl et al. 2004). Tout d'abord, l'échelle d'autocorrélation spatiale locale est dérivée d'un ajustement rapide (grossier) des résidus locaux basé sur une estimation itérativement croissante de ${}^m \varphi_{st}$ à une fonction d'autocorrélation de Matérn.

Pour traiter le temps de la même manière, il faudrait aussi déterminer les nœuds temporels et définir les échelles d'autocorrélation temporelle appropriées. En pratique, les données temporelles sont souvent rares et limitées dans les données d'enquête, de sorte que les données de toutes les périodes sont utilisées pour effectuer une mise à l'échelle brute, qui correspond essentiellement à une autocorrélation spatiale établie selon une moyenne temporelle. Une fois les limites approximatives du sous-domaine (soutien) estimées, les ${}^m \varphi_{st}$ sont modélisés comme une série de Fourier avec deux harmoniques, une interannuelle et une infra-annuelle (saisonniers) : f_m (interannuelle, saisonniers). Autrement dit, on n'utilise pas un modèle d'autocorrélation temporelle complet (covariance), mais plutôt un modèle fondé seulement sur un sous-ensemble des composantes à longueurs d'onde fixes :

$$\begin{aligned} {}^m \varphi_{st} &= f_m(\cdot) + \zeta_{st} \\ {}^m \zeta_{st} &\sim N(0, \zeta, \sigma_m^2), \end{aligned}$$

L'autocorrélation temporelle est donc portée par les processus temporels individuels à chaque point de référence spatial, et l'erreur ${}^\tau \sigma_m^2$ temporelle est la composante de la variance du modèle $f_m(\cdot)$, c'est-à-dire ${}^\tau \sigma_m^2 = \varphi \sigma_m^2 - \zeta \sigma_m^2$.

La fonction d'autocorrélation spatiale est paramétrée comme étant dérivée du processus gaussien moyen du sous-domaine avec une fonction de covariance de Matérn avec paramètres $\omega \theta_m = \{\phi_m \nu_m\}$ et une erreur type $\omega \sigma_m$ spatialement structurée variant dans le temps. Comme les données utilisées pour estimer la structure d'autocorrélation spatiale sont souvent rares, les données sont complétées par des prédictions temporelles du processus d'erreur résiduelle à chaque point de référence spatial (et désignées par un astérisque pour la notation). Ces processus résiduels augmentés temporellement sont modélisés spatialement pour chaque tranche temporelle φ_{st}^* comme la somme d'un **processus gaussien** spatial ${}^m \omega_{st}$ variant dans le temps et paramétré comme fonction de covariance spatiale de Matérn :

$$\omega \sigma_{mt}^2 \frac{1}{2^{\nu_{mt}-1} \Gamma(\nu_{mt})} (\sqrt{2\nu_{mt}} h / \phi_{mt})^{\nu_{mt}} K_{\nu_{mt}}(\sqrt{2\nu_{mt}} h / \phi_{mt})$$

avec erreur spatiale locale $\omega \sigma_{mt}$ et un processus d'erreur non structuré dans l'espace et dans le temps supposé dérivé d'un processus d'erreur normale avec une moyenne de zéro et une erreur $\sigma_{\varepsilon|m}$:

$$\begin{aligned} \varphi_{mt}^* &= \omega_{mt} + \varepsilon_{mt} \\ \omega_{mt} &\sim \text{GP}(0 \ C(\mathbf{s} \ \mathbf{s}'; \omega \boldsymbol{\theta}_{mt} = \{\nu_{mt} \ \phi_{mt} \ \omega \ \sigma_{mt}\})) \\ \varepsilon_{mt} &\sim \text{Normale}(0 \ \varepsilon \sigma_m^2), \end{aligned}$$

Cette approche représente un équilibre pratique entre le temps de calcul et la complexité/réalisme du modèle. Pour plus de rapidité, on utilise une convolution de Matérn basée sur la transformation rapide de Fourier. Cette solution met l'accent sur la prédiction à l'aide d'une mosaïque de solutions dans l'espace; la probabilité globale ou l'évaluation d'AIC pose un défi. À l'heure actuelle, le succès prédictif est le seul moyen d'évaluer l'utilité et, éventuellement, une certaine forme d'approches d'espérance-maximisation au moyen d'itérations si les vitesses de calcul s'améliorent, ce qui permettrait d'obtenir des estimations des paramètres moins biaisées.

MODÈLES HYBRIDES CARTÉSIENS-LAGRANGIENS

Les modèles spatiaux sans limites spatiales fixes (c.-à-d. les modèles *lagrangiens*) sont intuitifs et élégants. Toutefois, l'application pratique dans des situations pauvres en données entraîne des difficultés et des défis nécessitant le recours aux complexités du modèle qui peuvent s'intensifier de façon exponentielle lorsqu'on tente de rendre justice à la nature non stationnaire naturelle de ces processus spatiotemporels. Une autre approche consiste à traiter l'autocorrélation spatiotemporelle de manière discrète, mais locale. Cette approche commence par les unités spatiales de l'approche cartésienne, mais elle les élargit en permettant des interactions locales. La nature de ces interactions est simplement déterminée par la proximité (distance) ou la connectivité réelle (routes, sentiers, courants océaniques). Elle doit donc être déterminée avec soin. Cela permet de combiner l'accent du modèle *cartésien* sur les unités spatiales et l'évolution des domaines sous-jacents de la *variabilité des écosystèmes* des modèles *lagrangiens*. À mesure que la taille des unités spatiales diminue, les solutions de cette approche hybride convergeront, en théorie, avec celles des modèles *lagrangiens*, en raison du théorème de Hammersly-Clifford (voir ci-dessous). Ces modèles autorégressifs conditionnels peuvent être mis en œuvre sous forme bayésienne au moyen de [l'INLA](#). Nous utilisons la bibliothèque [carstm](#) pour faciliter la modélisation en agissant comme un simple point de départ de l'INLA qui gère les données requises pour ces modèles et les prévisions qui en résultent.

Modèles autorégressifs conditionnels spatiaux

Le modèle CAR est une extension du modèle de régression généralisée et donc, au départ, il est très semblable à la configuration des modèles *lagrangiens*. La région d'étude $D = D_{11} \dots D_{U1} \dots D_{UV}$ est un ensemble d'unités spatiales $i = 1 \dots U$ et d'unités temporelles $v = 1 \dots V$. Les observations \mathbf{Y} sur un tel ensemble D démontrent généralement une autocorrélation spatiale et temporelle, et peuvent être modélisées comme la même extension d'un modèle linéaire généralisé :

$$Y \sim f(\mu, \varepsilon \sigma^2)$$

$$g(\mu) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \varepsilon$$

Comme pour l'approche de la fonction d'autocorrélation, l'approche structurée CAR décompose également la structure d'erreur résiduelle en deux composantes aléatoires additives. Ces composantes sont une erreur spatiale structurée selon CAR ϕ et une erreur non spatiale (non structurée) *iid* ε :

$$g(\mu) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{O} + \phi + \varepsilon$$

La différence est que l'on accorde une attention particulière à la connectivité entre les unités spatiales. Cette connectivité entre dans l'estimation de ϕ et est représentée par convention par les éléments w_{ij} d'une matrice de proximités $W_{U \times U}$ de sorte que la connectivité est désignée par $w_{ij} = 1$, la non-connectivité est désignée par $w_{ij} = 0$ et l'autoconnectivité, c.-à-d. $\text{diag}(W)$, est désignée comme $w_{ii} = 0$. L'erreur ϕ d'autocorrélation spatiale résiduelle est traitée comme une variable aléatoire qui est censée avoir une *distribution conditionnelle locale* entre les treillis avoisinants $\{p(\phi_i | \phi_{i \sim j})\}$, à l'exclusion de soi $i \sim j$) qui se rapproche de la *distribution conjointe globale* $p(\phi_1 \phi_2 \dots \phi_U)$, sur la base du théorème de Hammersly-Clifford (Besag 1974, Besag et al. 1991, Leroux et al. 2000). La *distribution conditionnelle locale* (Banerjee et al. 2004) peut être exprimée sous forme de somme pondérée de la quantité aléatoire, où les poids sont dérivés de la *connectivité du quartier local* dans la matrice de proximités :

$$p(\phi_i | \phi_{i \sim j}, \tau) = N \left(\alpha \sum_{i \sim j}^K w_{ij} \phi_j, \tau^{-1} \right).$$

Ici, τ est un paramètre de précision et $\alpha \in [0 \ 1]$ contrôle la force de la dépendance/association spatiale, alors lorsque $\alpha = 0$, il n'y a pas de structure spatiale et $\alpha = 1$ indique un modèle autorégressif entièrement intrinsèque (conditionnel) : I(C)AR. Ce qui précède équivaut essentiellement à supposer une autocorrélation spatiale globale unique α , tempérée par la moyenne locale des erreurs spatiales pour trouver la structure spatiale approximative de la zone d'intérêt. Toute différence de comportement d'autocorrélation dans différents domaines doit donc être associée à des variations de covariables. Cette hypothèse n'est pas réaliste dans de nombreux systèmes, car les covariables peuvent ne pas exister correctement, ne pas être connues ou être en mesure d'exprimer ces différences entre les sous-zones, surtout lorsque des discontinuités réelles existent (voir stmv).

Sous l'hypothèse que ϕ est un champ aléatoire de Markov, la *distribution conjointe* de la variable aléatoire est, par l'entremise du lemme de Brook (Besag 1974) :

$$\phi \sim N(\mathbf{0}, Q^{-1}),$$

où,

$$Q = \tau D(I - \alpha B)$$

est la matrice de précision, et D est une matrice diagonale de même forme que W , et $\text{diag}(D) = d_{ii}$ = le nombre de voisins pour chaque US_{i_i} . En outre, $B = D^{-1}W$ est la matrice de proximité mise à l'échelle qui quantifie la force/l'influence relative de la connectivité pour chaque voisin; $\text{diag}(B) = b_{ii} = 0$. Voici la formule simplifiée :

$$Q = [\tau(D - \alpha W)]^{-1}.$$

Le paramètre de dépendance spatiale est habituellement difficile à estimer et est donc souvent présumé être $\alpha = 1$. Cela simplifie la matrice de précision pour :

$$Q = [\tau(D - W)]^{-1}.$$

Les ϕ sont non identifiables (l'ajout d'une constante à ϕ fait en sorte que la distribution conjointe demeure inchangée), ce qui en fait un *a priori* inapproprié. Autrement dit, la matrice de covariance est singulière et des contraintes supplémentaires sont nécessaires pour qu'elle soit traçable par calculs. Par convention, cette contrainte est qu'elle totalise zéro ($\sum_i \phi_i = 0$). Ce modèle est une simplification bien connue et fréquemment utilisée du modèle CAR, et il est connu sous le nom de I(C)AR, pour modèle autorégressif conditionnel intrinsèque (intrinsic conditional autoregressive). Lorsqu'il est appliqué à un processus de Poisson, il est également connu dans la littérature épidémiologique comme un modèle de convolution ou un modèle Besag-York-Mollie (BYM) (Besag *et al.* 1991).

L'établissement d'*a priori* sensibles sur la précision des effets aléatoires structurés et non structurés s'est avéré difficile, car ils dépendent de la connectivité de la zone d'intérêt (Bernardinelli 1995). Simpson *et al.* (2017) et Riebler *et al.* (2016), après Leroux *et al.* (2000), abordent cette question en paramétrant le problème avec ce qu'ils appellent le modèle bym2 dans l'INLA :

$$\psi = \psi(\sqrt{1 - \rho}\varepsilon^* + \sqrt{\rho}\phi^*)$$

où ψ est l'écart type global; $\rho \in [0 1]$ modélise l'équilibre relatif de l'hétérogénéité spatiale et non spatiale; $\varepsilon^* \sim N(0 1)$ est l'erreur aléatoire non structurée avec écart type fixe de 1, et la variable aléatoire ICAR est mise à l'échelle de sorte que $\text{Var}(\phi^* = \phi/v) \approx 1$. Ces hypothèses font en sorte que $\text{Var}(\varepsilon) \approx \text{Var}(\phi) \approx 1$, ce qui fait en sorte que ψ peut être interprété comme un écart type global. Les *a priori* peuvent alors être compris de façon plus conventionnelle et attribués comme étant des *a priori* avant l'écart type, comme une loi moitié-normale ou moitié-t ou exponentielle et un Bêta(0 5 0 5) pour ρ . Le modèle bym2 d'INLA utilise par défaut la distribution de Gumbel de type 2 pour ρ , qui réduit la distribution à zéro (Simpson *et al.* 2017). Pour le processus spatial (bym2), on utilise l'*a priori* PC recommandé (rho=0,5, alpha=0,5). Pour le processus AR1, on utilise l'*a priori* cor0 PC qui a un modèle « de base » de corrélation 0, c'est-à-dire aucune corrélation.

Comme l'abondance est généralement modale le long des gradients environnementaux, les covariables sont ici modélisées comme un processus de marche aléatoire de deuxième ordre (appelé « rw2 » dans INLA), elle est discrétisée le long de ses quantiles (9 points de discrétisation). Cela permet une représentation fonctionnelle lisse de la relation locale des covariables qui, dans le cas d'un indice i espacé uniformément, est :

$$\Delta^2 x_i = (x_i - x_{i+1}) - (x_{i+1} - x_{i+2}) \sim N(0, \tau^{-1})$$

Il s'agit d'une perspective locale qui s'apparente philosophiquement à l'approche CAR. L'*a priori* PC recommandé pour la précision est utilisé. Il s'agit d'une fonction exponentielle qui se transforme en un modèle « de base » de 0 précision. Bien qu'il s'agisse d'additifs dans l'espace de liaison logarithmique, ils sont multiplicatifs dans les modèles de Poisson qui suivent.

Modèles spatiotemporels autorégressifs conditionnels

La variation temporelle du processus focal d'intérêt peut également être traitée comme un processus autorégressif. L'argument, semblable au cas spatial, est qu'il y a une continuité dans le temps pour de nombreux processus, et donc une autocorrélation temporelle inhérente. Si des covariables existent et présentent cette autocorrélation temporelle, l'autocorrélation temporelle est induite et introduite mécaniquement dans le modèle. Toute autocorrélation non prise en compte ou induite doit toutefois être explicitement prise en compte pour assurer que $\varepsilon \stackrel{iid}{\sim} f(\cdot)$. Dans la littérature sur les séries chronologiques, ces modèles sont connus sous le nom de modèles de type AR(I)MA (AutoRegressive Integrated Moving Average – moyenne mobile intégrée autoregressive), souvent utilisés lorsque les modèles déterministes sont inexistantes ou difficiles à paramétrer.

Dans le contexte de l'évaluation des stocks halieutiques, le modèle d'évaluation des stocks est habituellement un modèle de séries chronologiques soumis à des contraintes liées aux hypothèses de processus de taux biologiques, comme la mortalité et la croissance. La pratique actuelle consiste à regrouper toutes les unités spatiales et à les traiter comme un seul processus de série chronologique intégré. Il s'agit bien sûr d'une solution provisoire utilisée pour éviter les difficultés liées à la complexité du modèle et aux limites historiques de la capacité de calcul. Lorsque l'ensemble ou la plupart des unités spatiales agissent de manière homogène, cette approximation peut suffire pour exprimer le processus focal d'intérêt. Ce n'est toutefois pas toujours le cas, surtout dans le contexte océanographique, puisque la variabilité temporelle et spatiale se produit à toutes les échelles.

À l'heure actuelle, conformément à l'accent mis sur les processus autorégressifs simples et opérationnels dans les modèles CAR et sur l'obtention de solutions viables sur le plan informatique, nous étendons le modèle CAR par l'ajout simple d'un processus temporel AR(1).

COMPARAISON DES RÉSULTATS

Le choix des unités spatiales utilisées dans les modèles fondés sur des unités spatiales est souple. Les régions de gestion (figure 1) auraient pu être utilisées, mais elles auraient été trop grossières pour permettre de repérer les agrégations locales et les écarts dans la répartition, des renseignements utiles pour évaluer l'état des stocks. Une autre solution consiste à utiliser les unités spatiales du poisson de fond (figure 4), mais ce sont vraiment des unités spatiales fondées sur la profondeur, qui ne tiennent pas compte des autres covariables environnementales. Bon nombre de ces dernières unités sont également assez grandes et ne correspondent donc pas entièrement à l'échelle spatiale des modèles d'abondance du crabe des neiges. D'autres solutions consistent à modifier l'un ou l'autre des éléments ci-dessus pour réduire la taille des polygones de plus grande taille au besoin ou même simplement à utiliser des polygones de Voronoï ou des algorithmes de maillage (p. ex., actuellement mis en œuvre dans l'approche « carstm » et le modèle INLA).

Dans le relevé sur le crabe des neiges, le plan initial était fondé sur des treillis (10 min X 10 min) et nous avons donc adopté ici une structure simple en treillis, où les unités spatiales sont définies comme des unités de 25 km X 25 km, légèrement plus grandes que la grille originale de 10 min X 10 min (figure 10). Il a été déterminé qu'il s'agissait d'un équilibre raisonnable entre le temps de calcul et le RAM requis pour modéliser (important dans les modèles spatiotemporels CAR plus complexes) par rapport à des solutions raisonnablement représentatives suffisantes pour décrire les tendances spatiales. Évidemment, s'attendre à ce qu'une zone de 625 km² soit homogène en ce qui a trait à la variabilité des écosystèmes et les densités de crabe des neiges demeure une hypothèse audacieuse (dans le simple cas sans

covariable). Les unités spatiales ont été limitées à celles pour lesquelles au moins trois relevés ont été effectués depuis 1996. Les limites du polygone ont également été resserrées pour suivre les isobathes de 0 et de 350 m, les limites d'exploitation des chaluts de relevé et de la pêche commerciale.

MODÈLES NON SPATIAUX ET NON TEMPORELS

Pour simplifier la comparaison avec les modèles fondés sur le modèle CAR, dans la présente section, nous utilisons les mêmes unités spatiales que celles qui ont été élaborées pour les calculs fondés sur le modèle CAR (figure 10). Comme ces représentations non spatiales et non temporelles sont minimalement paramétrées, cela signifie essentiellement qu'elles reposent sur des hypothèses audacieuses sur les données : qu'il n'existe pas d'autocorrélations et de biais spatiotemporel induit par la variabilité environnementale. On suppose plutôt que chaque échantillon est *parfaitement représentatif* de l'unité spatiale d'où il est dérivé et, surtout, qu'il n'existe aucune relation entre les unités spatiales. Bien qu'irréaliste, ce plan *aléatoire stratifié* constitue une base simple à partir de laquelle on peut comparer la pertinence d'un *plan expérimental* pour représenter le processus spatial, d'autant plus qu'il s'agit d'une supposition très fréquente.

La supposition de répartition la plus courante est une supposition gaussienne/normale :

$$Y_{st} \sim \text{Normale}(\mu_{st} \quad \psi \sigma_{st}^2)$$

$$\mu_{st} = \beta_{st} + \psi_{st}$$

Les variations de la biomasse dans ce modèle, que nous appellerons « **factoriel croisé gaussien (biomasse)** » (figure 11; A2.1) ne représentent que des estimations agrégées *lorsqu'il y a des échantillons*. Lorsque les unités spatiales ne sont pas échantillonnées en raison des limites logistiques (2017) ou de la perte d'un bateau (2014), un biais spatial sera induit, car elles doivent être présumées ponctuelles pour avoir une certaine valeur (habituellement zéro). Par conséquent, les faibles valeurs des années 1999 et 2000 sont attribuables à la faible portée du relevé. De même, lorsque les visites ont lieu à des moments différents de l'année (avant 2004 et après 2004), un biais temporel (saisonnier) peut être induit. Tous ces facteurs rendent l'utilisation d'une approche factorielle croisée insensée pour les comparaisons interannuelles et ne sont présentés ici que pour leur valeur heuristique. Dans la section qui suit, nous nous concentrerons particulièrement sur l'année 2017, car en raison de problèmes logistiques, le relevé n'a pas été terminé et certaines régions n'ont pas été échantillonnées (zones extracôtières; figure 11).

Étant donné que la répartition d'abondance animale est généralement mise à l'échelle logarithmique, la forme de modèle suivante peut être utile :

$$Y_{st} \sim \text{Normale}(\mu_{st} \quad \psi \sigma_{st}^2)$$

$$\ln(\mu_{st}) = \beta_{st} + \psi_{st}$$

Malheureusement, bien que l'information à valeur positive s'adapte raisonnablement bien, il existe de nombreux emplacements de densité de biomasse à valeur nulle qui doivent nécessairement être ignorés. Cela est dû au fait que $\log(\text{zéro})$ n'est pas défini. Dans l'écologie numérique, il était (et est toujours) courant d'ajouter une petite constante à ces zéros, mais cela n'est pas conseillé, car l'ampleur de cette constante peut encore avoir des effets importants et biaiser énormément les estimations. Ces emplacements à valeur nulle sont toutefois informatifs : certains sont attribuables à une absence réelle et d'autres à des erreurs d'observation comme les limites de détection des engins d'échantillonnage dans des situations de faible nombre ou les relevés incomplets. Les emplacements à valeur nulle devraient être

traités comme faisant partie d'un modèle de deuxième étape qui définit les emplacements à valeur nulle. Encore une fois, un modèle apparemment simple, mais avec un doublement du nombre de paramètres serait nécessaire. Ce modèle n'est pas examiné numériquement dans le présent rapport, bien qu'une approche similaire soit utilisée dans l'analyse fondée sur l'approche stmv. La complexité de calcul supplémentaire est contraire à notre désir d'obtenir une solution opérationnelle simple pour l'intégration des données.

Une alternative plus fonctionnelle est une supposition de Poisson qui fonctionne selon des densités numériques plutôt que des densités de biomasse. C'est aussi le plus simple de ces modèles, car il n'y a qu'un paramètre d'intensité (par opposition à une moyenne et à une variance dans le gaussien et le log-normal) :

$$Y_{st} \sim \text{Poisson}(\mu_{st})$$
$$\ln(\mu_{st}) = \beta_{0|st} + O_{st} + \psi_{st}$$

où O sont les mesures compensatoires (zone de relevé). De plus, les valeurs nulles sont permises et les réalisations sont des nombres entiers et non des fractions d'un individu. Si la *synthèse* de la pêche fonctionne selon l'abondance numérique, c'est un avantage. Si la biomasse est nécessaire, les nombres doivent être convertis en biomasse au moyen de certaines suppositions liées au poids moyen des individus dans un échantillon, ce qui est facilement réalisable puisque le poids et les nombres sont enregistrés. Si elle est absente, elle peut aussi être imputée. Par conséquent, nous privilégions la supposition de répartition de Poisson.

Les solutions temporelles agrégées dérivées de cette supposition de répartition de Poisson que nous appellerons un modèle « **factoriel croisé** » (figure 12a; A2.1) sont en effet très semblables à celles d'une opération directe sur la biomasse et d'une supposition de répartition gaussienne (« **factoriel croisé gaussien [biomasse]**; figure 11), bien que la variabilité des estimations soit réduite. Les différences sont plus claires dans la représentation spatiale où l'agrégation spatiale est plus groupée (moins diffuse) dans le modèle de Poisson par rapport au modèle gaussien. Encore une fois, le manque de suppositions sur la façon de traiter les emplacements non échantillonnés rend les comparaisons interannuelles inutiles.

Il convient de souligner à nouveau que dans les modèles factoriels, lorsqu'il y a des unités spatiales qui ne sont pas échantillonnées, il n'y a aucun moyen d'estimer l'abondance dans ces zones, car il n'y a pas de suppositions pour le permettre. Ceci contraste avec les méthodes fondées sur le modèle CAR et les méthodes d'autocorrélation qui permettent une telle imputation/interpolation, ce à quoi nous allons maintenant nous intéresser.

MODÈLES D'AUTOCORRÉLATION CONTINUE

Avant de poursuivre dans cette section, le lecteur est averti que les résultats de cette approche ne sont pas disponibles actuellement. Le temps de calcul requis pour réaliser cette approche est maintenant passé de deux à trois mois. Ceci n'est pas opérationnel et c'est donc la raison de ce nouveau cadre. Ce qui est présenté dans cette section est à jour en ce qui concerne les parties spatiales seulement, mais les résultats spatiotemporels proviennent d'analyses antérieures et ne sont mentionnés ici que pour fournir un contexte.

Pour des raisons pragmatiques, à savoir le temps et les ressources de calcul, les prédictions sont discrétisées en résolution de 1 km X 1 km et en résolution temporelle de 10 unités infra-annuelles. Comme l'approche est continue, ces discrétisations sont arbitraires et facilement modifiables en fonction des ressources de calcul disponibles et des exigences du problème. Pour les évaluations du crabe des neiges, cette résolution a été considérée comme un bon équilibre entre le gain d'information, la résolution de la variabilité de l'écosystème et les

exigences en matière de données (RAM) et de vitesse (CPU), les coûts et la viabilité dans le délai des évaluations (environ un mois entre la disponibilité des données et l'évaluation).

Chacune des covariables utilisées dans cette approche nécessite également une modélisation et une prédiction pour obtenir une base espace-temps cohérente. Pour le crabe des neiges, ces covariables comprennent : la profondeur, la pente du fond, la courbure du fond, la taille des grains du substrat, les températures au fond et la composition taxinomique du fond. Les deux premières ont de longues échelles temporelles qui sont par rapport à l'échelle temporelle du crabe des neiges et peuvent donc être traitées comme un problème purement spatial.

La **bathymétrie** (profondeur; m ; = -élévation) est une caractéristique très informative en ce sens qu'elle détermine les niveaux de lumière ambiante, la complexité/rugosité de la surface, la stabilité hydrodynamique et la stabilité environnementale globale. Comme il n'y a pas de covariables, il s'agit d'un modèle extrêmement simple. Toutefois, la densité des données est extrêmement élevée et variable. Les sources sont variées, y compris les relevés du crabe des neiges, les relevés du poisson de fond, le SHC, etc. Ici, nous le modélisons comme un processus log-normal localisé :

$$\begin{aligned} \log(Y_s) &= F(\text{décalage constant}) + \varphi_s \\ \varphi_s &\sim \text{Normale}(0 \ \sigma_\varphi^2) \\ \varphi_{s|m} &= \omega_{s|m} + \varepsilon_{s|m} \\ \omega_{s|m} &\sim \text{GP}(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_m = \{\nu_m \ \phi_m \ \sigma_m\})) \\ \varepsilon_{s|m} &\sim \text{Normale}(0 \ \sigma_{\varepsilon|m}^2), \end{aligned}$$

Le processus d'erreur résiduelle $\varphi_{s|m}$ est facilement décomposé en un processus spatial local $\omega_{s|m}$ et une erreur locale non structurée $\varepsilon_{s|m}$. Une convolution de Matérn basée sur la transformation rapide de Fourier est utilisée pour exprimer le processus spatial d'amélioration de la vitesse de calcul. La réduction des données est utilisée pour résoudre les variations à courte échelle et les détails granulaires; cette réduction dépend de la densité des données (la similarité de 90 % à 50 % du niveau d'autocorrélation spatiale a tendance à bien fonctionner). Fait important, on observe que même dans une variable simple comme la profondeur, la moyenne et l'écart type varient de façon complexe. Autrement dit, la bathymétrie n'est **pas** stationnaire de premier et de deuxième ordre et que, dans les modèles spatiotemporels plus complexes, les modèles séparables simples sont probablement inappropriés (figure 12b).

Comme dans le cas de la bathymétrie, la taille des grains du substrat (mm) peut être considérée comme un modèle purement spatial, car elle varie également selon les échelles de temps géologiques. Bien sûr, des changements catastrophiques rapides peuvent se produire et se produisent effectivement, mais ils peuvent quand même être considérés de façon statique, car nous ne nous concentrons pas sur les processus géologiques. La taille des grains est en fin de compte une mesure indirecte du type de substrat (boue, sable, gravier, roche, etc.) et peut donc être utile pour les organismes benthiques, démersaux et endofauniques.

Malheureusement, la seule donnée disponible actuellement est une surface (sur)lissée (Kostylev et Hannah 2007). Certaines données ont été ajoutées à partir des relevés du crabe des neiges. Elle est également modélisée comme un processus log-normal, problème de convolution de covariance de Matérn fondé sur la transformation rapide de Fourier :

$$\begin{aligned}
\log(Y_s) &= F(\text{profondeur, pente, courbure}) + \varphi_s \\
\varphi_s &\sim \text{Normale}(0 \ \sigma_\varphi^2) \\
\varphi_{s|m} &= \omega_{s|m} + \varepsilon_{s|m} \\
\omega_{s|m} &\sim \text{GP}(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_m = \{v_m \ \phi_m \ \sigma_m\})) \\
\varepsilon_{s|m} &\sim \text{Normale}(0 \ \sigma_{\varepsilon|m}^2),
\end{aligned}$$

La taille des grains du substrat est clairement associée à la profondeur (figures 13, 14; A2.2). Toutefois, étonnamment, il s'agit d'une relation modale. D'après les études observationnelles, on s'attendait à une baisse de l'abondance avec la taille des grains. La taille des grains dans la fourchette de 50 mm (=exp(4)) à 250 mm (=exp(5,5)) semble être optimale. Indépendamment de la cause, cette relation nous permet donc d'extrapoler bien au-delà du domaine spatial des données relatives au substrat, en empruntant le contenu informationnel de la *variabilité des écosystèmes* (dans ce cas, les variables liées à la profondeur).

La **température** est un modulateur fondamental du métabolisme, de la croissance, de la reproduction, de la répartition et de l'abondance des prédateurs et des proies, de l'incidence des maladies, de la composition taxinomique, etc. Les températures au fond, en particulier, font l'objet d'une attention particulière en raison de leur pertinence pour le crabe des neiges et sont modélisées comme un processus spatiotemporel hiérarchique et « non séparable ». Comme elles présentent des variations de fréquence élevée, une certaine complexité supplémentaire est nécessaire pour modéliser leurs tendances spatiotemporelles. Ici, les effets temporels sont emboîtés dans les sous-domaines spatiaux S_m . Pour ce faire, on utilise une approche biharmonique où les sinusoides sont utilisés pour représenter des variations interannuelles et infra-annuelles des séries chronologiques à chaque nœud statistique en utilisant des données considérées comme faisant partie du sous-domaine local. Les prédictions dans l'espace sont ensuite estimées comme un problème purement spatial sous la supposition désormais juste de stationnarité locale, pour chaque tranche temporelle. Autrement dit, il s'agit d'un modèle (hiérarchique) non séparable.

Le modèle global des covariables est simplement un modèle d'interception avec un lien d'identité de sorte que $\varphi_{s \ t}$ sont centrés sur zéro. Les données sur la salinité ou la densité de l'eau peuvent être utilisées pour délimiter les masses d'eau et leurs origines, mais ces données ne sont pas d'une qualité, d'une densité et d'une couverture suffisantes pour être assez informatives et justifier la charge de calcul supplémentaire (à l'heure actuelle). Les erreurs résiduelles sont modélisées localement dans chaque sous-domaine sous forme de série chronologique avec deux harmoniques de Fourier dans le temps (une composante interannuelle et une composante infra-annuelle/saisonnnière). Des termes lisses supplémentaires pénalisés de fonction spline à plaque mince pour la profondeur et l'emplacement locaux sont utilisés pour résoudre les tendances spatiales locales et le repliement vers un troisième ordre ou moins (par contraction). Les prédictions temporelles à chaque point de référence spatial sont ensuite utilisées pour « accroître » la modélisation des processus spatiaux $\varphi_{s \ t|m}^*$ qui sont traités indépendamment pour chaque tranche temporelle comme un **processus gaussien**. L'autocorrélation temporelle n'est donc portée qu'indirectement par les processus temporels individuels centrés à chaque point de référence spatial. Pour des calculs plus rapides, une méthode de convolution fondée sur la transformation rapide de Fourier est utilisée pour approximer le processus gaussien spatial. La spécification du modèle est donc la suivante :

$$\begin{aligned}
Y_{s t} &= F(\text{identité}) + \varphi_{s t} \\
\varphi_{s t} &\sim \text{Normale}(0 \sigma_{\varphi}^2) \\
\varphi_{s t|m} &= f_m(\text{interannuelle, saisonnière, ordonnée, abscisse, profondeur}) + \zeta_{s t|m} \\
\zeta_{s t|m} &\sim \text{Normale}(0 \sigma_{\zeta|m}^2) \\
\varphi_{s t|m}^* &= \omega_{s t|m} + \varepsilon_{s t|m} \\
\omega_{s t|m} &\sim \text{GP}\left(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_{t|m} = \{v_{t|m} \ \phi_{t|m} \ \sigma_{t|m}\})\right) \\
\varepsilon_{s t|m} &\sim \text{Normale}(0 \sigma_{\varepsilon|m}^2),
\end{aligned}$$

Les données sur la température proviennent d'un certain nombre de sources différentes, dont le relevé du crabe des neiges, le relevé du poisson de fond, le relevé du PMZA, le relevé de la FSRS, le relevé du pétoncle et de nombreux autres échantillons opportunistes conservés et gracieusement fournis par Roger Petitpas (DSO, MPO).

Les résultats ne sont pas présentés dans ce rapport, car le temps de calcul est significatif (>2 mois) compte tenu des ressources de calcul disponibles. Les représentations proviennent de versions antérieures du modèle (figure 15).

La composition taxinomique du fond est modélisée de la même manière que la température au fond. La composition taxinomique est simplement les deux premières décompositions propres (analyse des composantes principales des vecteurs propres) de la matrice de corrélation de l'abondance des espèces (figure 16).

$$\begin{aligned}
Y_{s t} &= F(\text{identité}) + \varphi_{s t} \\
\varphi_{s t} &\sim \text{Normale}(0 \sigma_{\varphi}^2) \\
\varphi_{s t|m} &= f_m(\text{interannuelle, saisonnière, ordonnée, abscisse, profondeur, température}) + \zeta_{s t|m} \\
\zeta_{s t|m} &\sim \text{Normale}(0 \sigma_{\zeta|m}^2) \\
\varphi_{s t|m}^* &= \omega_{s t|m} + \varepsilon_{s t|m} \\
\omega_{s t|m} &\sim \text{GP}\left(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_{t|m} = \{v_{t|m} \ \phi_{t|m} \ \sigma_{t|m}\})\right) \\
\varepsilon_{s t|m} &\sim \text{Normale}(0 \sigma_{\varepsilon|m}^2),
\end{aligned}$$

Les données de la présente analyse sont tirées des relevés du crabe des neiges et des relevés du poisson de fond. La période s'étend de 1999 à aujourd'hui, car avant cette période, dans les relevés du poisson de fond, l'identification des espèces était incohérente pour la plupart des invertébrés. Le premier axe de représentation est principalement associé aux températures tandis que le deuxième est principalement associé aux profondeurs.

Le processus d'estimation du crabe des neiges utilise les covariables ci-dessus ainsi que les covariables interpolées pour obtenir le même soutien spatial de l'évaluation du crabe des neiges aux fins de prédiction. Cela est nécessaire pour préciser les prédictions de l'abondance du crabe des neiges et de son habitat tout en évitant les problèmes de biais attribuables au repliement.

Des covariables supplémentaires qui expriment l'état de l'écosystème à un moment et à un endroit donnés (indicateurs) sont utiles pour délimiter les processus spatiotemporels structurés par rapport à ceux qui sont aléatoires. La formulation de leur modèle est similaire en ce sens qu'ils suivent une structure de modèle similaire dont les effets temporels sont emboîtés dans les sous-domaines spatiaux et l'utilisation de fonctions de lien dans un cadre de modèle linéaire généralisé ou un modèle additif généralisé où les covariables utilisées pour modéliser ces indicateurs s'appuient sur des prédictions spatiales de la profondeur et de la taille des grains du substrat et sur les dérivés spatiaux des premières (pente et courbure). Le processus d'erreur

spatiotemporelle est modélisé localement dans chaque sous-domaine comme un modèle espace-temps non séparable, en utilisant des covariables variant en fonction du temps et liées aux variations de température de fond et aux statistiques connexes.

Pour l'estimation des préférences en matière d'habitat et la création de cartes de répartition des espèces (figure 17), les données sur la présence-absence (Y), sont supposées provenir d'un processus binomial de Bernoulli avec une fonction de lien logit $g(\cdot)$:

$$\begin{aligned}
 \text{logit}(Y_{s t}) &= F(\text{profondeur, pente, courbure, taille des grains du substrat}) + \varphi_{s t} \\
 \varphi_{s t} &\sim \text{Normale}(0 \sigma_{\varphi}^2) \\
 \varphi_{s t|m} &= f_m(\text{indicateurs d'écosystème}) + \zeta_{s t|m} \\
 \zeta_{s t|m} &\sim \text{Normale}(0 \sigma_{\zeta|m}^2) \\
 \varphi_{s t|m}^* &= \omega_{s t|m} + \varepsilon_{s t|m} \\
 \omega_{s t|m} &\sim \text{GP}\left(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_{t|m} = \{v_{t|m} \ \phi_{t|m} \ \sigma_{t|m}\})\right) \\
 \varepsilon_{s t|m} &\sim \text{Normale}(0 \ \sigma_{\varepsilon|m}^2),
 \end{aligned}$$

Pour l'estimation de l'abondance (figure 18), les données positives Y sont supposées provenir d'un processus log-normal :

$$\begin{aligned}
 \log(Y_{s t}) &= F(\text{profondeur, pente, courbure, taille des grains du substrat}) + \varphi_{s t} \\
 \varphi_{s t} &\sim \text{Normale}(0 \ \sigma_{\varphi}^2) \\
 \varphi_{s t|m} &= f_m(\text{indicateurs d'écosystème}) + \zeta_{s t|m} \\
 \zeta_{s t|m} &\sim \text{Normale}(0 \ \sigma_{\zeta|m}^2) \\
 \varphi_{s t|m}^* &= \omega_{s t|m} + \varepsilon_{s t|m} \\
 \omega_{s t|m} &\sim \text{GP}\left(0 \ C(\mathbf{s} \ \mathbf{s}'; \boldsymbol{\theta}_{t|m} = \{v_{t|m} \ \phi_{t|m} \ \sigma_{t|m}\})\right) \\
 \varepsilon_{s t|m} &\sim \text{Normale}(0 \ \sigma_{\varepsilon|m}^2),
 \end{aligned}$$

Les résultats plus actuels de cette approche analytique, bien qu'à la fine pointe de la technologie, ne sont pas disponibles en raison de la puissance et du temps de calcul, du moins jusqu'à ce que la puissance de calcul augmente suffisamment.

MODÈLES AUTORÉGRESSIFS CONDITIONNELS

Les résultats fondés sur le modèle CAR sont basés sur une discrétisation de l'espace à une échelle de 25 km X 25 km et sur 10 unités infra-annuelles dans le temps (figure 10). La structure du quartier a été déterminée en fonction d'une règle de contiguïté de la Reine (tout contact le long de toutes les frontières, y compris les diagonales). Des unités plus petites auraient pu être choisies pour aider à déterminer les processus locaux, mais cela aurait rendu les calculs plus lents et aurait également risqué d'imputer des données pour une très grande proportion d'unités spatiales où l'échantillonnage serait insuffisant et où des voisins ne seraient pas échantillonnés. À l'inverse, des unités plus grandes entraîneraient moins d'imputations, mais les unités spatiales ne seraient pas en mesure de résoudre les caractéristiques spatiales importantes. Quoi qu'il en soit, différentes formulations de polygone peuvent être explorées à l'avenir pour voir si des avantages supplémentaires en matière de précision et d'exactitude pourraient être réalisés par rapport aux coûts du temps de calcul et de l'instabilité de la solution.

Modèle Poisson-CAR

Alors que nous favorisons le modèle de répartition de Poisson de l'abondance numérique (figure 19), nous lui ajoutons une seule autocorrélation spatiale CAR comme effet aléatoire pour

voir l'effet incrémentiel (A2.3). Nous appellerons cela le modèle « **simple à effets mixtes** », car l'espace est traité comme un effet aléatoire et l'année comme un effet fixe. Cette supposition permet l'estimation dans les unités spatiales non échantillonnées. La tendance temporelle est semblable à celle du modèle qui n'est pas CAR, mais les grandeurs agrégées sont plus élevées comme on peut s'y attendre par rapport aux modèles à effets fixes. Dans la représentation spatiale, nous pouvons constater que cela est dû à une valeur attendue plutôt optimiste (extrapolation/imputation) des unités spatiales voisines non échantillonnées. En outre, comme l'échantillonnage n'est pas entièrement représentatif du domaine et qu'il est en fait préférentiel (figure 6), ces estimations devraient être positivement biaisées.

Variabilité des écosystèmes sur le plan des variables statiques

Il faut tenir compte de la variabilité de l'écosystème pour tenter d'ajuster cet échantillonnage préférentiel. Cela se fait facilement dans un schéma d'inférence et de prédiction fondé sur le modèle CAR en tant que covariables (et l'année est toujours traitée comme un facteur, de sorte qu'il s'agirait d'un modèle à « effets mixtes »). Toutefois, ces facteurs doivent être exprimés selon le même espace et les mêmes échelles de temps que le processus focal d'intérêt. Cela peut être accompli de manière *ponctuelle* à partir des données disponibles. Cependant, de telles tentatives présenteraient les mêmes problèmes de base liés à la réduction de l'exactitude et de la précision que toute approche *cartésienne*. Il serait prudent de les modéliser. Bien qu'il soit possible d'utiliser des champs modélisés selon l'approche stmv, il est plus cohérent, du point de vue des suppositions et de la méthodologie, d'utiliser des méthodes basées sur le modèle CAR pour ces champs de covariables. Il convient toutefois de souligner qu'en raison de la très grande taille de ces unités spatiales par rapport à la discrétisation dans le stmv, certaines variables comme la courbure et la pente du fond ne sont pas aussi informatives et ne sont donc pas calculées. À l'heure actuelle, seul un nombre limité de variables écosystémiques sont disponibles pour l'analyse systémique. Nous présentons ici un résumé de certains de ces résultats.

La profondeur du fond – en raison de sa corrélation avec un grand nombre de processus biologiques, physiques et chimiques. La solution fondée sur le modèle CAR résume les profondeurs à chaque unité spatiale, en fonction des suppositions d'autocorrélation et d'erreur non structurée (figure 20). Les variations spatiales sont conformes à la bathymétrie connue dans la région (p. ex., figure 12b). La représentation est moins satisfaisante dans les zones où les profondeurs changent rapidement sur de courtes distances (p. ex., banc de Misaine, zones près du rivage) où les unités spatiales, en raison de leur grande taille, n'ont pas été en mesure de résoudre la variabilité, ce qui a entraîné une réduction de la moyenne à des grandeurs intermédiaires (figure 20). Il convient de noter que les composantes de corrélation spatiale dominent l'erreur non structurée, ce qui indique que les corrélations entre les unités spatiales sont élevées (figure 20, au bas). En effet, la répartition a posteriori du facteur de proportionnalité ρ qui indique l'importance relative des erreurs spatiales est beaucoup plus élevée que celle des erreurs non structurées : mode de 0,996 et intervalle de crédibilité bayésien (ICB) à 95 % : 0,922; 0,999.

La **taille des grains du substrat** est également un facteur important pour les organismes benthiques et démersaux, car ils indiquent les processus géologiques et hydrodynamiques prédominants dans une région. Pour la zone d'intérêt, la taille des grains est associée à la profondeur (figure 21). Les processus d'erreurs autocorrélées spatiales dominent les erreurs non structurées, avec un mode a posteriori ρ de 0,959 et un ICB à 95 % de (0,867, 0,995).

Il est à noter qu'il existe certaines différences dans le rapport entre la taille des grains et la profondeur dans les approches du stmv (figure 14) et du modèle CAR (figure 22). Bien que la taille maximale des grains soit atteinte près des mêmes profondeurs (50 m à 100 m), la

modalité est moins prononcée. Cela démontre l'effet de la résolution spatiale et de l'agrégation des données (souvent appelée le « problème de l'unité spatiale cartographique ») qui réduit la capacité de l'approche CAR à résoudre les variations à petite échelle, particulièrement dans les zones à topographie variable, proches du rivage, de 10 m à 30 m.

Nous appellerons l'ajout de ces covariables statiques au modèle « simple à effets mixtes » un modèle « statique à effets mixtes » (figure 23). Nous observons que les effets aléatoires spatiaux dominaient toujours les erreurs non spatiales dans le modèle « statique à effets mixtes », à $\rho = 0,89$ (A2.4), une augmentation marginale par rapport à $\rho = 0,84$ dans le modèle « simple à effets mixtes » (A2.3). Selon divers critères théoriques d'information (DIC, WAIC, CPO), il s'agit d'une amélioration par rapport au modèle « simple à effets mixtes ». Cependant, cela s'est fait au prix d'une augmentation du nombre de paramètres efficaces, associés aux paramétrages des covariables. Dans l'ensemble, les grandeurs sont réduites par un déplacement des densités vers des zones considérées comme associées positivement à la densité numérique, en fonction de l'information sur les covariables, principalement la profondeur (figure 24). La taille des grains du substrat a un effet plus faible, ainsi qu'un effet et une relation plus variables, en partie parce qu'elle est elle-même fortement associée à la profondeur. Toutefois, une augmentation marginale de la densité est suggérée avec une augmentation de la taille des grains, ce qui contredit surtout les observations empiriques (figure 25). Cela s'explique probablement par l'observation empirique de la taille des grains du substrat confondu avec la profondeur (Hooper 1986) et pourrait ne pas avoir été ajustée pour cet effet.

Variabilité des écosystèmes sur le plan des variables dynamiques

La **température** est un facteur biologique omniprésent qui influence le crabe des neiges à de nombreux égards, que ce soit indirectement par la disponibilité de la nourriture, la prévalence des prédateurs et la prévalence des maladies, ou directement par la croissance, la reproduction et les processus métaboliques de base. Pour ajouter ce facteur, il faut également le modéliser à un état cohérent. Ici, nous nous concentrons sur les températures au fond, car elles dominent la majeure partie de la durée de vie du crabe des neiges. La variabilité spatiotemporelle élevée de la température au fond peut être modélisée simplement comme un modèle CAR-AR1 non séparable (A2.5; figures 26 et 27). Afin d'assurer la résolubilité de calcul, nous emboîtons les effets CAR dans chaque année et utilisons un processus AR1 pour modéliser les variations annuelles. De plus, nous ajoutons un processus AR1 additif (séparable) pour la saisonnalité et un processus RW2 lié à la profondeur (marche aléatoire de deuxième ordre, exponentielle) pour tenir compte de la variabilité de la température liée à la profondeur. Les autocorrélations spatiales représentaient 42,7 % de la variabilité (A2.5; figure 27). Les autocorrélations temporelles à l'échelle annuelle étaient de 0,841, alors qu'à l'échelle infra-annuelle, l'autocorrélation était de 0,725 (A2.5).

La relation entre les températures de fond et les autres covariables environnementales est complexe. Par rapport à la profondeur du fond, une relation de type pot-et-poignée est observée avec des températures minimales régulièrement observées à des profondeurs intermédiaires entre 70 m et 90 m (figure 28). Ceci est surtout associé au courant de la Nouvelle-Écosse, une masse d'eau froide qui circule le long de ces profondeurs, près du littoral.

En ce qui concerne la saisonnalité dans le domaine de l'évaluation du crabe des neiges, les températures moyennes au fond sont les plus basses de janvier à mai, puis augmentent de façon significative d'environ 2 °C en moyenne pour le reste de l'année (figure 29).

La variabilité interannuelle est limitée à la durée de la série du crabe des neiges, bien qu'elle ne doive pas l'être. Au cours de cette période, les variations moyennes ont été de la même ampleur que les variations saisonnières (environ 2 °C; figure 30). De 2010 à 2018, les

températures moyennes au fond ont été plus élevées, par rapport à la période plus froide de 2001 à 2009.

La **composition taxinomique** à un endroit et à un moment donnés est un marqueur très informatif des conditions environnementales ambiantes. Les deux premiers vecteurs propres (dominants) résument l'information relative à la température et à la profondeur d'un emplacement donné (figures 31 et 32) et sont chacun soumis à la même structure de modèle que celle utilisée pour la température. Cependant, la température est maintenant une covariable supplémentaire. Pour le premier axe (PCA1), environ 51,9 % de l'erreur peut être attribuée à la structure spatiale (A2.6), tandis que pour le deuxième axe, ce pourcentage était de 53,8 % (A2.7).

Les prévisions globales étaient semblables aux résultats fondés sur l'approche stmv, mais l'agrégation a pour effet d'estomper spatialement les tendances (figures 33, 34). Les interprétations des axes sont similaires, l'ACP1 étant principalement un schéma d'association d'espèces en fonction de la température et l'ACP2 étant associé à la profondeur (tableaux 6 et 7). L'autocorrélation temporelle aux échelles annuelles est élevée à 0,919 pour l'ACP1 et à 0,914 pour l'ACP2. Les autocorrélations infra-annuelles étaient inférieures à 0,055 et à 0,601, respectivement. Cette dernière était plus forte puisque les associations liées à la « profondeur » changeaient moins que les associations liées à la « température » qui fluctuaient davantage aux échelles saisonnières. Il convient de noter que la série chronologique de l'ACP1 et de l'ACP2 indique que les assemblages d'espèces ont fortement changé depuis 2015, passant d'un assemblage dominé par l'eau froide à la fin des années 2000 à un assemblage dominé par l'eau plus chaude. Les changements dans l'ACP2 sont également semblables, les assemblages d'eau plus profonde étant devenus plus importants au début de 2000 et passant à une dominance d'assemblages d'eaux moins profondes depuis 2010. Ces derniers peuvent être associés à des changements de pycnocline dus à des changements de température, de salinité et à l'hydrodynamique du mélange des masses d'eau froide à faible salinité avec les eaux chaudes et salées du Gulf Stream.

La densité du **crabe des neiges** est facilement modélisée comme un processus spatiotemporel en ajoutant des covariables spatiales et spatiotemporelles. Nous appellerons ce modèle « **dynamique à effets mixtes** » (A2.8). L'effet spatial représente 76 % de la variabilité supplémentaire de Poisson. La série chronologique agrégée (figure 35) indique une réduction des grandeurs globales, car les covariables aident à peaufiner l'imputation trop optimiste dans les emplacements non échantillonnés.

La façon dont l'année est traitée est un élément important. Dans les modèles à effets mixtes, elle est traitée comme un facteur. En réalité, il existe une autocorrélation temporelle. Cela peut être représenté par un processus AR1 additif qui signifie un modèle spatiotemporel « **séparable** » (figure 36; A2.9). Le modèle suggère une autocorrélation significative de 0,86 entre les années. L'effet spatial représente 85 % de la variation non représentée par la loi de Poisson. L'ajustement global est encore une fois amélioré selon la mesure DIC. Toutefois, l'effet sur la série chronologique agrégée de la biomasse et des prédictions spatiales est minime (figure 36). Les limites de l'utilisation d'un seul effet CAR pour toutes les années semblent avoir été atteintes.

La prochaine question raisonnable est donc de savoir si l'espace et le temps ne sont pas séparables. Une façon de procéder est d'avoir un effet CAR pour chaque année (« **espace/temps non séparable** ».) L'alternative consiste à avoir un seul effet CAR comportant plusieurs processus AR1 pour chaque unité spatiale (« **espace|temps non séparable** »). Il s'agit d'une espérance raisonnable, car il n'y a aucune raison valable de penser

que les relations spatiales dans le nombre total de crabes des neiges persistent d'une année à l'autre.

La mise en œuvre est encore une fois assez simple. Tout d'abord, sans aucune covariable, le modèle « **simple non séparable** » (A2.10 s'avère être similaire aux autres modèles « simples » sans la variabilité de l'écosystème comme covariable : les grandeurs sont élevées, mais la série chronologique est plus lisse. Et comme pour les autres modèles « simples », l'absence de variabilité des écosystèmes donne lieu à une prédiction optimiste de l'abondance dans les unités spatiales sans échantillons (figure 37).

L'ajout de covariables qui intègrent la variabilité des écosystèmes permet de peaufiner ces prédictions trop optimistes (modèle « **espace|temps non séparable** » – A2.11; figure 38). L'effet spatial est significatif à 66 % de la variation non représentée par la loi de Poisson et une force d'autocorrélation interannuelle de 0,80. L'autre, « **espace|temps non séparable** » (non montré), affichait un rendement semblable en fonction de divers critères de qualité de l'ajustement (tableau 1).

La façon dont les covariables interagissent avec les densités numériques du crabe des neiges au moyen de ce modèle « **espace|temps non séparable** » est la suivante (figure 39). La profondeur a une très forte relation avec les densités maximales du crabe des neiges qui se trouvent à 150 m et plus. De même, la taille des grains du substrat a un rapport positif avec la densité. Cela va à l'encontre des attentes fondées sur la littérature quant à la préférence pour la boue et laisse entendre qu'il s'agissait d'une observation confondue. Une densité accrue peut être liée à la complexité et à l'hétérogénéité de l'habitat et constitue une nouvelle observation. Les relations avec les températures sont reconnues comme importantes et la plage de 1°C à 4°C semble être associée aux densités les plus élevées. Toutefois, on constate une persistance à des températures allant jusqu'à 10 °C, mais de manière plus variable. La saisonnalité observée est probablement associée à une augmentation des effectifs due à la mue au printemps ou en été, puis à une diminution subséquente due à la mortalité naturelle et par pêche en automne et en hiver. La relation avec PCA1 est frappante dans la mesure où les assemblages d'espèces sont associés à un gradient de température. La présence d'assemblages en eau froide est un puissant prédicteur de densités plus élevées de crabe des neiges. PCA2 est associé à un gradient de profondeur et suggère que les profondeurs intermédiaires sont les plus favorables aux densités élevées de crabe des neiges. Il convient de souligner que ces relations sont légèrement différentes de celles dérivées des modèles plus simples (p. ex. les modèles « statiques » des figures 19 et 23).

Dans les séries chronologiques agrégées, tous les modèles affichent une tendance générale similaire (figure 40) : le crabe des neiges a fluctué sur au moins un cycle dont la périodicité est d'environ 12 ans, ce qui correspond à la longévité approximative d'un crabe des neiges mâle. La question est de savoir de combien étaient ces amplitudes et quelle était leur taille relative. Trois grands groupes de solutions sont observées à la figure 40 :

1. celles qui ignorent les données sur l'écosystème comme les modèles « * simples » et les modèles « factoriels croisés » (sans covariables environnementales);
2. celles qui intègrent des caractéristiques environnementales statiques seulement (profondeur, substrat) – estimations de magnitude intermédiaire;
3. celles qui incluent des données sur l'écosystème variant dans le temps — estimations de magnitude faible.

La méthodologie *non spatiale et non temporelle* (cartésienne) de la « **biomasse factorielle croisée (gaussienne)** » se rapproche le plus d'une application naïve d'un modèle (spatial) « stratifié aléatoire » à un processus spatiotemporel (figure 40; ligne épaisse verte foncée). Bien

entendu, comme nous l'avons déjà mentionné, cela n'est pas utile en raison des nombreux cas d'échantillonnage incomplet qui rendent les comparaisons interannuelles impossibles. Pour que les comparaisons temporelles soient possibles, des hypothèses permettant de traiter les processus spatiaux ou temporels sont nécessaires.

L'application d'effets aléatoires CAR ou AR1 facilite la prédiction dans des zones sans échantillonnage (contrairement aux modèles factoriels) et très utiles dans le contexte d'un programme d'évaluation qui sont généralement sujettes à des défaillances logistiques. Toutefois, cette prédiction peut être trop optimiste lorsqu'elle est effectuée avec des modèles simples et il convient d'être prudent. L'ajout de covariables informatives qui indexent des mesures pertinentes de la variabilité de l'écosystème aide à moduler ces prédictions à des limites plus raisonnables. Cela laisse entendre également que même les prédictions de notre meilleur modèle surestiment probablement l'état réel, car les covariables qui représentent la variabilité de l'écosystème sont encore limitées et assez rudimentaires.

Les hypothèses de séparabilité de l'autocorrélation spatiale et temporelle sont répandues dans la littérature appliquée. Ici, nous notons que leurs prédictions sont semblables sur le plan qualitatif aux modèles non séparables, avec des intervalles de crédibilité à 95 % qui se chevauchent. Toutefois, les variations sont suffisamment informatives (fondées sur les mesures DIC, WAIC et CPO) pour suggérer que l'utilisation du modèle plus complexe serait avantageuse.

Fondamentalement, ces résultats suggèrent que l'estimation $\int y_{st} ds$ au moyen d'un schéma d'agrégation/d'intégration non séparable, spatiotemporel et basé sur un modèle de variabilité de l'écosystème est recommandée. De toute évidence, la pratique assez courante consistant à utiliser un modèle purement non spatial et non temporel ou même un schéma purement spatial pour représenter un processus spatiotemporel peut donner lieu à des estimations des paramètres biaisées. Ce résultat est-il surprenant? Pas vraiment.

SYNTHÈSE

La synthèse des données sur la pêche fonctionne sur la biomasse agrégée ($\int y_{st} ds$ ou plus souvent $\int Y_{st} ds$) au moyen d'un modèle de pêche. Dans l'évaluation du crabe des neiges, il s'agit d'un modèle (logistique) de dynamique de la [biomasse](#) (implanté en langage [Julia](#)).avec les [paramètres](#) connexes précisés.

L'application de ce modèle, peut-être le modèle espace-état le plus simple possible, est simple et permet d'équilibrer la variabilité temporelle avec les prises de pêche et les contraintes biologiques sur le taux d'augmentation intrinsèque r (figure 41), la capacité de charge K (figure 42) et la capturabilité q (figure 43).

En supposant que toutes les prises ont été comptabilisées correctement, le modèle de dynamique de la biomasse suggère que la biomasse de la composante exploitable est faible dans toutes les zones de gestion (figure 44). Toutes les zones suivent les taux d'exploitation cibles (20 %) et dépassent F_{RMD} occasionnellement seulement (figure 45).

Pour ce qui est de l'emplacement dans le cadre des règles de contrôle des prises, une zone se trouvait dans chaque secteur (figure 46).

CONCLUSIONS

Les méthodes présentées ici représentent une solution unifiée, viable et cohérente pour modéliser de manière opérationnelle l'abondance du crabe des neiges et la variabilité de l'écosystème dans lequel il vit et, par extension, de nombreuses autres espèces et leur niche écologique qui nécessitent une évaluation dans la région des Maritimes en raison des changements rapides de l'écosystème et du climat ainsi que des changements rapides dans les modes d'exploitation par l'homme et la perturbation de leur écosystème.

CODA

Modèles d'habitat (aussi appelés « modèles de répartition des espèces »)

Le modèle « **espace/temps non séparable** » est une forme de modèle opérationnel viable pour estimer l'abondance. Toutefois, l'utilité nécessite toujours un échantillonnage raisonnablement cohérent de l'organisme d'intérêt. Pour de nombreuses espèces, il n'existe pas d'ensemble cohérent d'échantillons. Par exemple, les relevés du poisson de fond (« écosystème ») ont systématiquement ignoré les espèces non commerciales au milieu des années 1980 et presque tous les invertébrés jusqu'en 1999. Dans les relevés du crabe des neiges, les espèces autres que le crabe des neiges ont été enregistrées de manière systématique après 2005 seulement, et en raison de la taille des mailles des engins, le petit crabe des neiges est peu capturé. Dans de telles situations, les estimations de l'abondance relative pourraient être non fiables, mais une délimitation de l'utilisation de l'habitat peut tout de même être estimée en modifiant les hypothèses de répartition du modèle linéaire généralisé de base en un cas binomial : $Y \sim \text{Binomial}(\eta, \theta)$, où $\ln(\theta/(1 - \theta)) = \mathbf{x}^T \boldsymbol{\beta} + \mathbf{F} + \boldsymbol{\Psi}$, et η est le vecteur du nombre d'essais avec θ le vecteur des probabilités de réussite dans chaque essai. En utilisant une telle hypothèse de répartition et en appliquant toutes les données disponibles, nous pouvons délimiter les préférences d'habitat du petit crabe des neiges immature (largeur de carapace < 50 mm) et ainsi estimer les changements spatiaux et temporels de cette importante composante de la population (A2.12; figure 47).

Comme on peut le constater, les principales zones de « **reproduction** » se trouvent dans des environnements près de la rive, peu profonds et légèrement plus chauds avec un assemblage d'espèces d'eau plus chaude (figure 48). Pour favoriser la viabilité à long terme de la pêche, une protection effective de ces zones contre les perturbations humaines doit faire l'objet des discussions approfondies. Une forte mortalité naturelle due à la prédation est probable dans ces zones également. La gestion écosystémique suggère que l'abondance des prédateurs (comme le flétan et la morue) dans ces zones doit être surveillée de près et que la pêche ciblée des prédateurs dans ces zones pourrait devoir être discutée avec d'autres secteurs de la pêche.

De plus, la même analyse de l'habitat pour la composante exploitable indique une diminution importante de l'habitat de cette composante depuis 2010 (figure 47). En revanche, les femelles matures ont des préférences en matière d'habitat légèrement différentes (eaux plus chaudes et moins profondes que les grands mâles), ce qui se traduit par une probabilité globale plus faible de les observer dans la région des Maritimes (figure 49). Le pic de 2007 et 2008 est associé à des températures au fond fraîches dans des eaux moins profondes (75 m à 150 m) près du rivage.

Bien que ces évaluations de l'habitat et de l'abondance soient extrêmement simplistes (p. ex. par rapport à l'approche stmv), leur application systématique permet néanmoins d'établir des comparaisons significatives de l'utilisation interspécifique de l'espace et des exigences ou sensibilités des écosystèmes en général, applications qui sont facilement adaptées aux

scénarios de changement climatique pronostiques. Il s'agit d'étapes modestes et simples, mais nécessaires aux fins d'une gestion des pêches véritablement écosystémique.

TABLEAUX

Tableau 1. Résumé des différentes représentations modélisées de l'abondance numérique du crabe des neiges dans la région des Maritimes dans le présent rapport. Toutes les représentations reposent sur une hypothèse de répartition de Poisson et sont modélisées avec l'INLA, à l'exception du modèle « factoriel croisé gaussien » qui applique une hypothèse gaussienne directement sur les densités de biomasse à l'aide du MLG. Ici, « croisé » signifie que seuls les termes d'interaction sont utilisés, sans effets marginaux (principaux). La notation « | » indique « groupé par », comme on l'utilise couramment pour la modélisation à effets mixtes. Tiret (-) = non pertinent.

Modèle	Unité spatiale	Année	Saison	Variabilité environnementale (rw2)	Logarithme du rapport de vraisemblance	DIC	WAIC	CPO
Factoriel croisé gaussien (biomasse)	fixe	fixe	-	-	-	-	-	-
Factoriel croisé	fixe	fixe	-	-	34 659	-	-	34 590
Covariable factorielle	fixe	fixe	ar1	profondeur+substrat+temp+PCA1+PCA2	30 409	-	-	122 636
Simple à effets mixtes	bym2	fixe	-	-	36 105	71 454	73 156	36 458
Statique à effets mixtes	bym2	fixe	-	profondeur+substrat	33 426	65 926	67 663	33 701
Dynamique à effets mixtes	bym2	fixe	-	profondeur+substrat+temp+PCA1+PCA2	30 115	59 302	61 066	30 368
Simple séparable	bym2	ar1	-	-	36 045	71 449	73 144	36 464
Séparable	bym2	ar1	ar1	profondeur+substrat+temp+PCA1+PCA2	30 052	59 301	61 054	30 368
Simple non séparable	bym2 année	ar1	ar1	-	26 684	53 557	62 148	32 952
Espace temps non séparable pas PCA	bym2 année	ar1	-	profondeur+substrat+température	23 713	47 310	55 052	29 009
Temps espace non séparable	bym2	ar1 AU	ar1	profondeur+substrat+temp+PCA1+PCA2	30 571	44 476	50 380	26 577
Espace temps non séparable	bym2 année	ar1	ar1	profondeur+substrat+temp+PCA1+PCA2	22 092	44 295	50 308	26 940

FIGURES

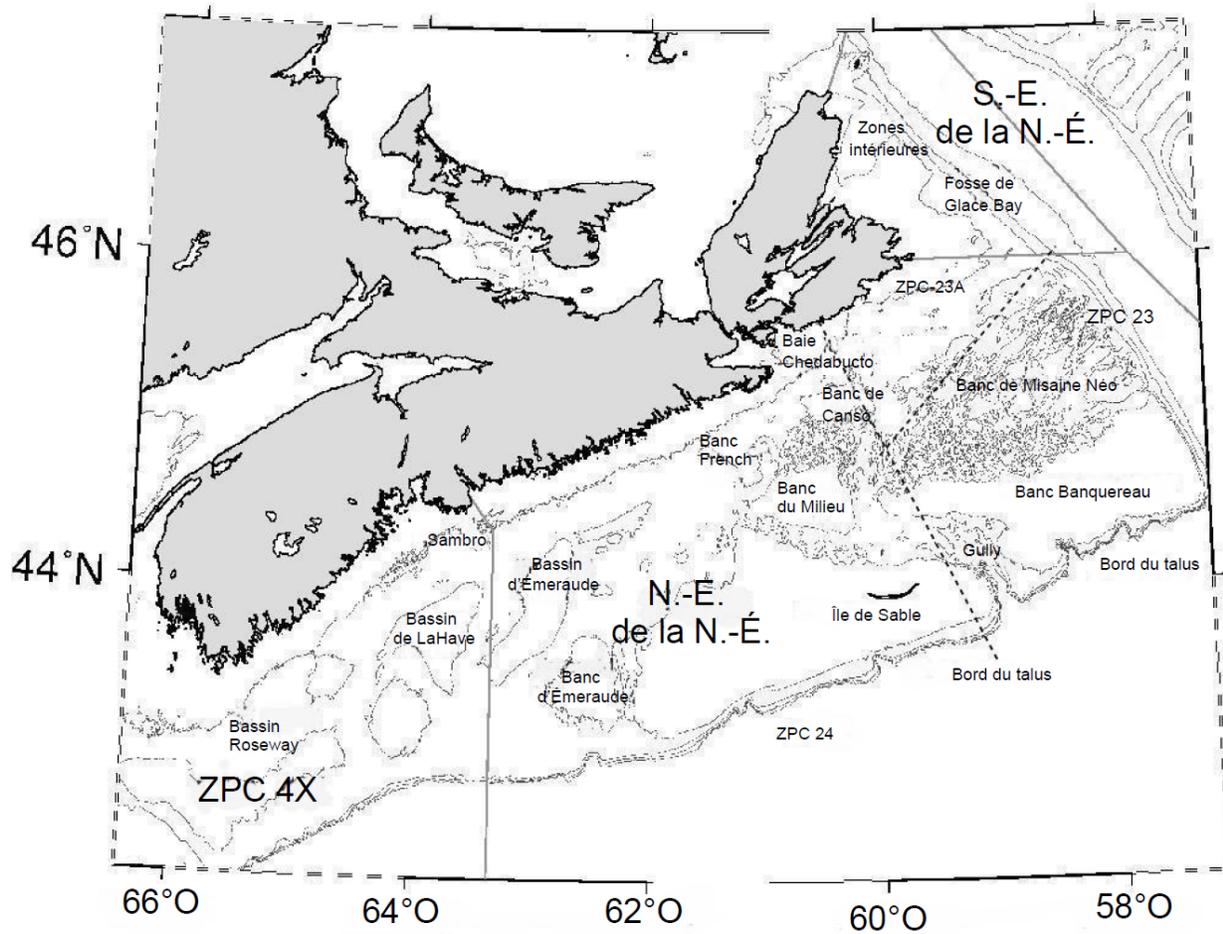


Figure 1. Zones de pêche du crabe des neiges dans la région des Maritimes.

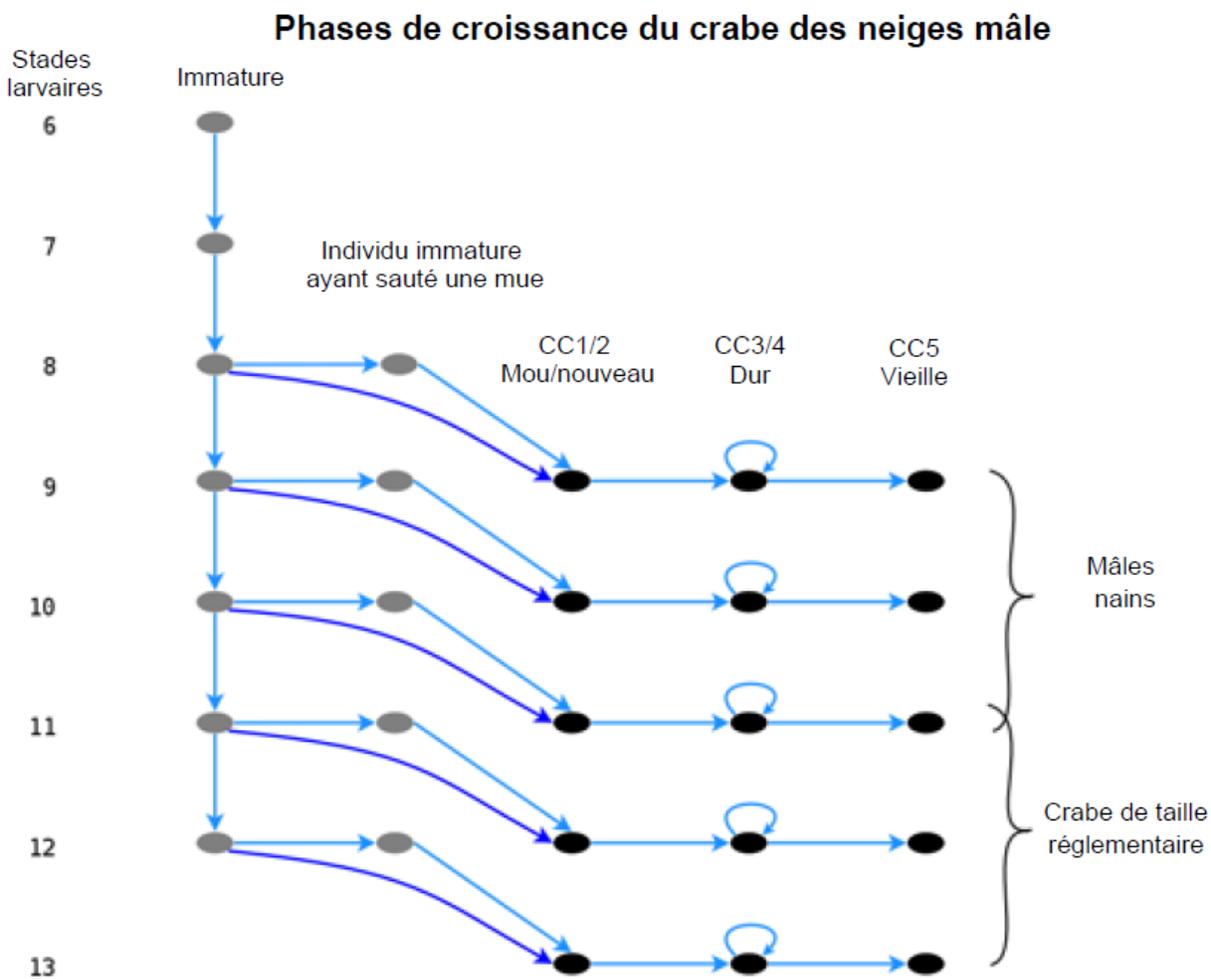
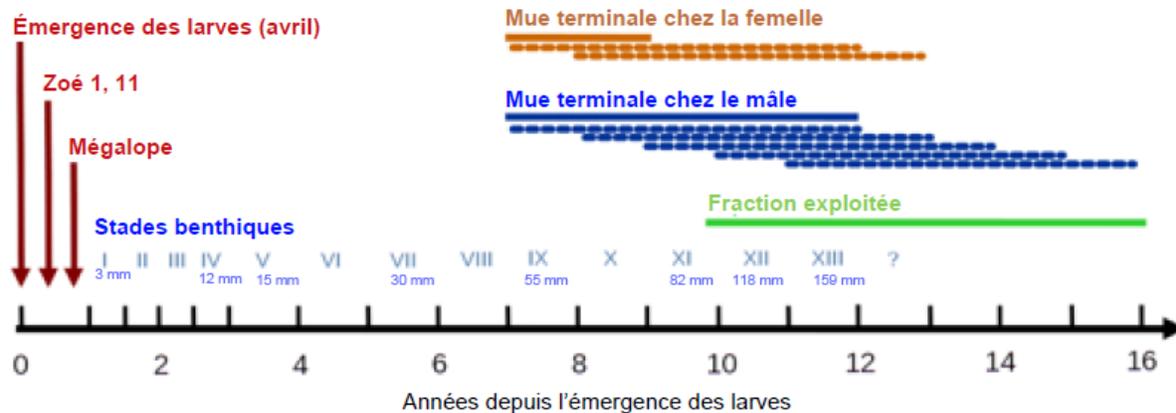


Figure 2. Modèles du cycle biologique du crabe des neiges. Haut : chronologie des principaux stades du cycle biologique du crabe des neiges. Bas : phases de croissance de la composante mâle et chemins de décision vers la maturité et la mue terminale.

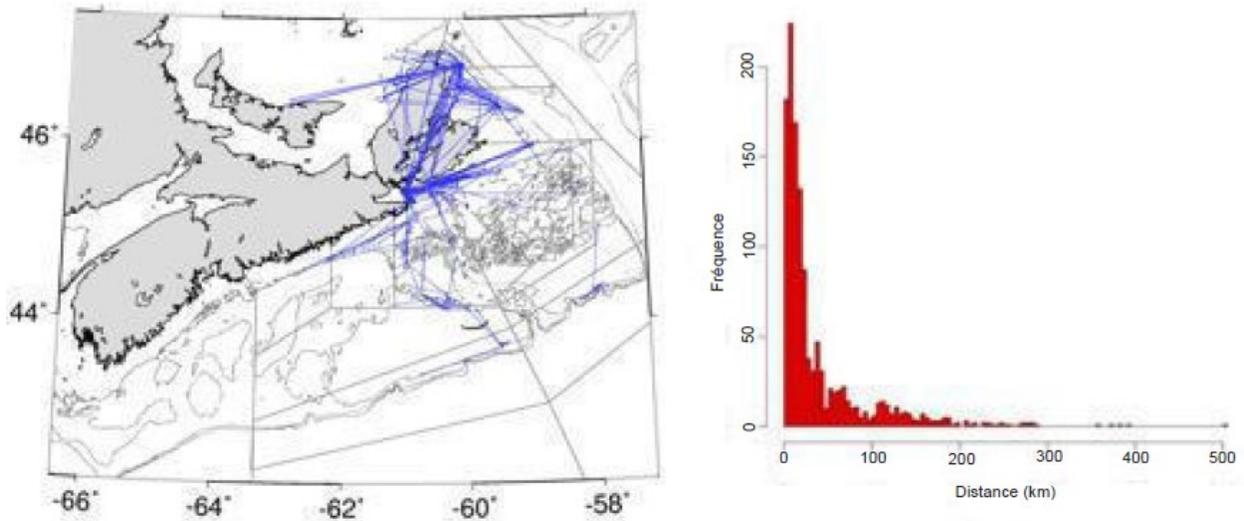


Figure 3. Déplacement du crabe des neiges et distances parcourues

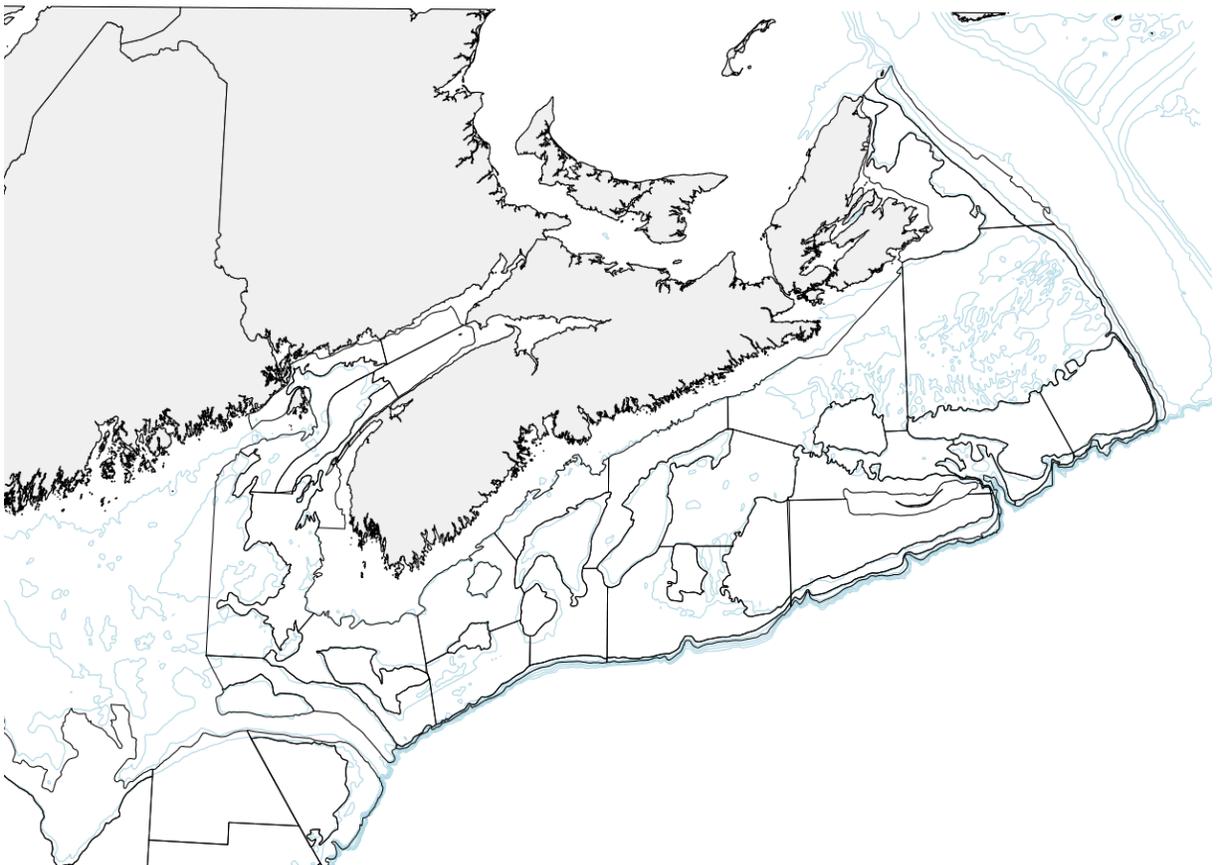


Figure 4. Exemple d'unités spatiales utilisées dans la région des Maritimes de l'Atlantique de l'est du Canada (lignes noires) pour l'échantillonnage du relevé écosystème (anciennement appelé relevé du poisson de fond) aux fins de l'évaluation de l'abondance des stocks de nombreuses espèces. Les isobathes (bleu pâle) superposées montrent que la plupart de ces unités spatiales suivent de fait des contours bathymétriques simples.

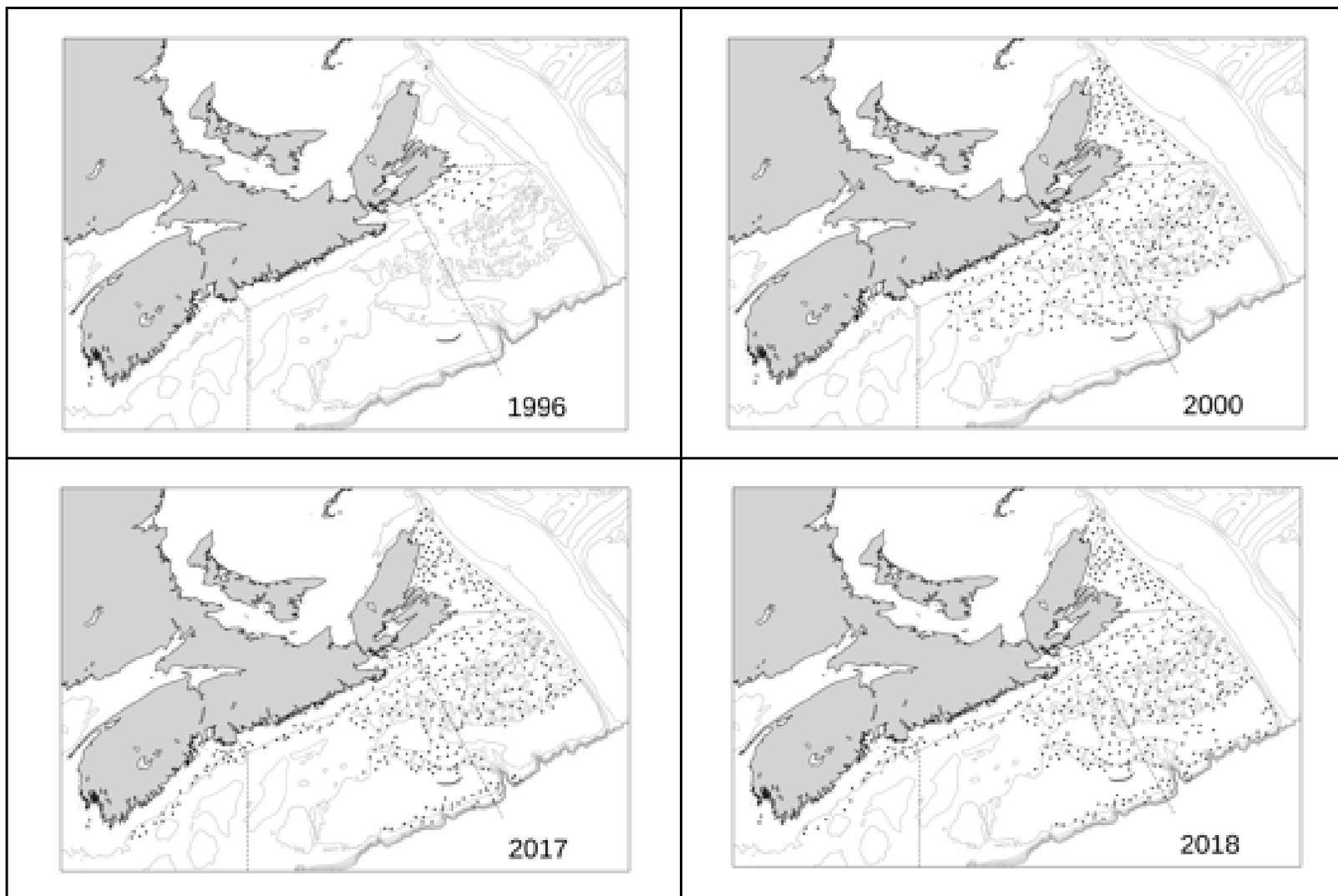


Figure 5. Variabilité des lieux d'échantillonnage dans le relevé du crabe des neiges.

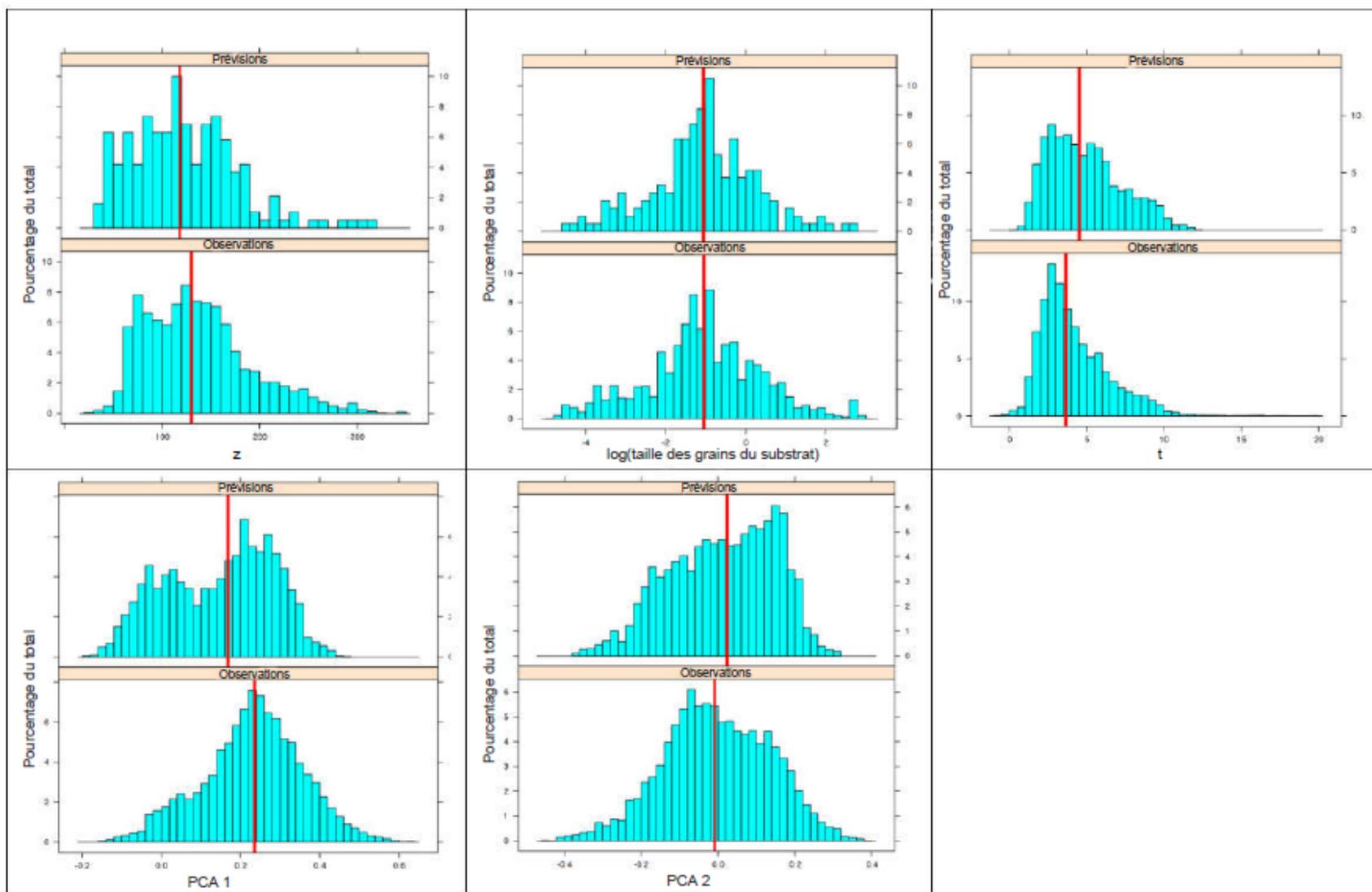


Figure 6. Biais d'échantillonnage dans la variabilité de l'écosystème entre les lieux étudiés (« observations ») et le domaine global (« prédictions ») pour la profondeur (z), la taille des grains du substrat, la température (t) et la composition de l'espèce (pca1, pca2).

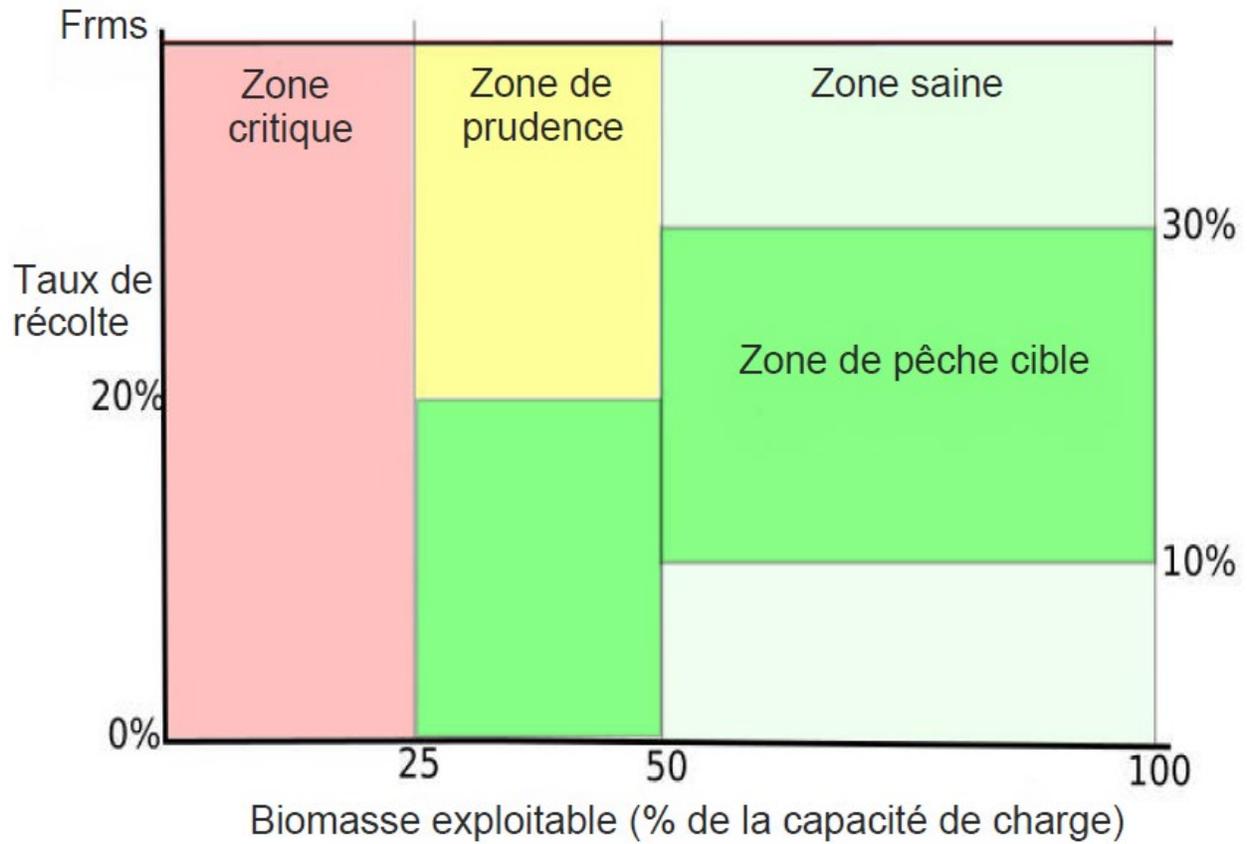


Figure 7. Stratégie de contrôle des prises s'inscrivant dans le contexte de l'approche préventive pour le crabe des neiges.

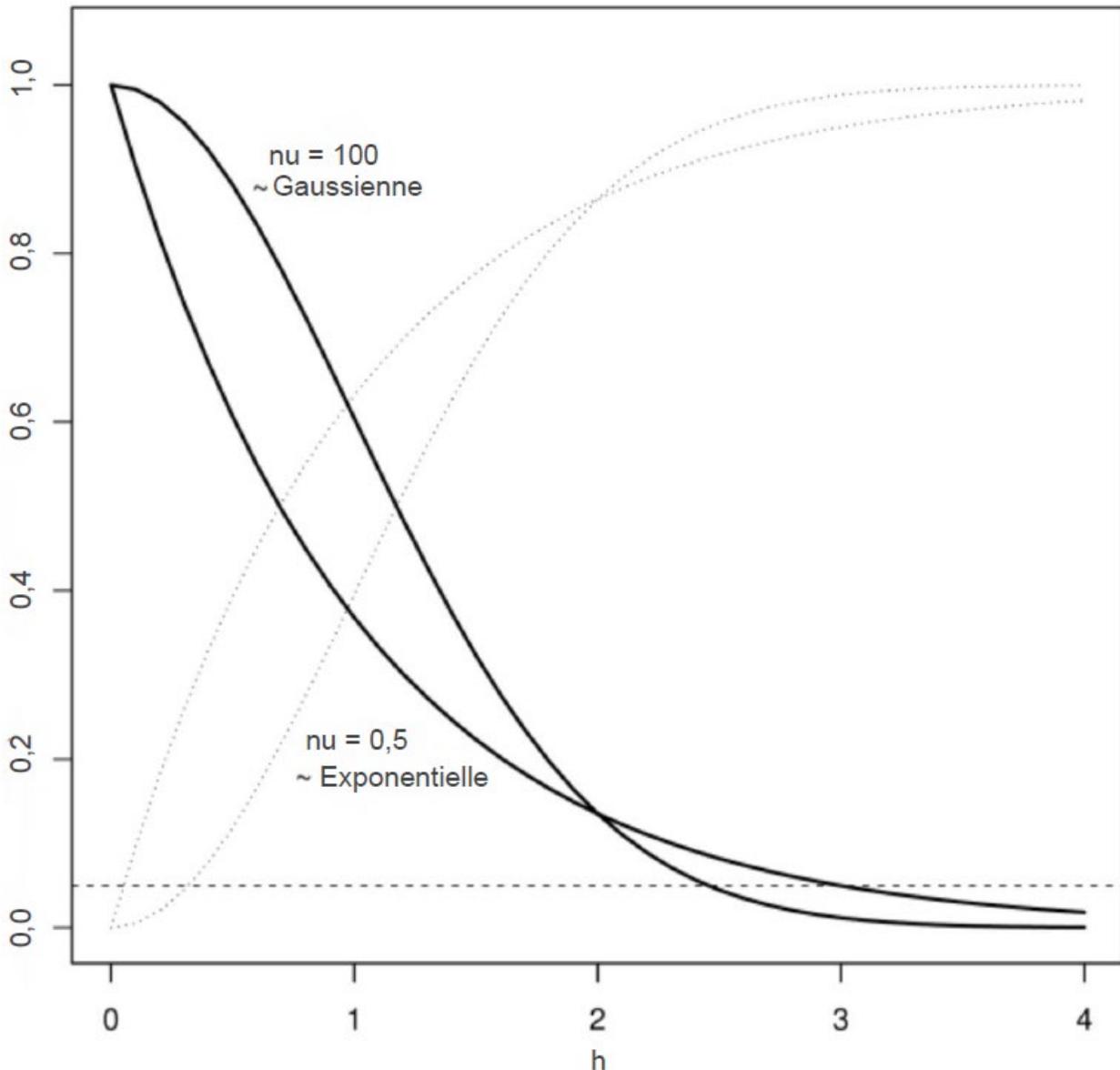


Figure 8. Fonction d'autocorrélation de Matérn, $\rho(h) = C(h)/C(0)$, la fonction de covariance $C(h)$ mise à l'échelle par la variance totale $C(0)$, pour deux valeurs de ν (lignes foncées). À mesure que ν augmente ($\nu = 100$), la valeur s'approche de la courbe gaussienne (courbe foncée supérieure à gauche), tandis qu'à des valeurs plus petites, ($\nu = 0.5$) la courbe est exponentielle (courbe foncée inférieure à gauche). Cette flexibilité a fait de cette fonction un choix populaire en géostatistique. Les semivariogrammes associés (mis à l'échelle de la variance unitaire) $\gamma(h)$ sont représentés par des lignes pointillées pâles. L'échelle spatiale est définie de manière heuristique comme la distance h à laquelle l'autocorrélation tombe à une valeur faible (ligne en tirets horizontale). La semivariance (également appelée semivariogramme), $\gamma(h)$, est plus couramment utilisée dans la littérature géostatistique et correspond simplement à la fonction de covariance $C(h)$ reflétée sur l'axe horizontal de la variance globale $C(0)$ telle que $\gamma(h) = C(0) - C(h) = \frac{1}{2} \text{Var}[Y_s - Y_{s'}] = \sigma^2 [1 - \rho(h)]$.

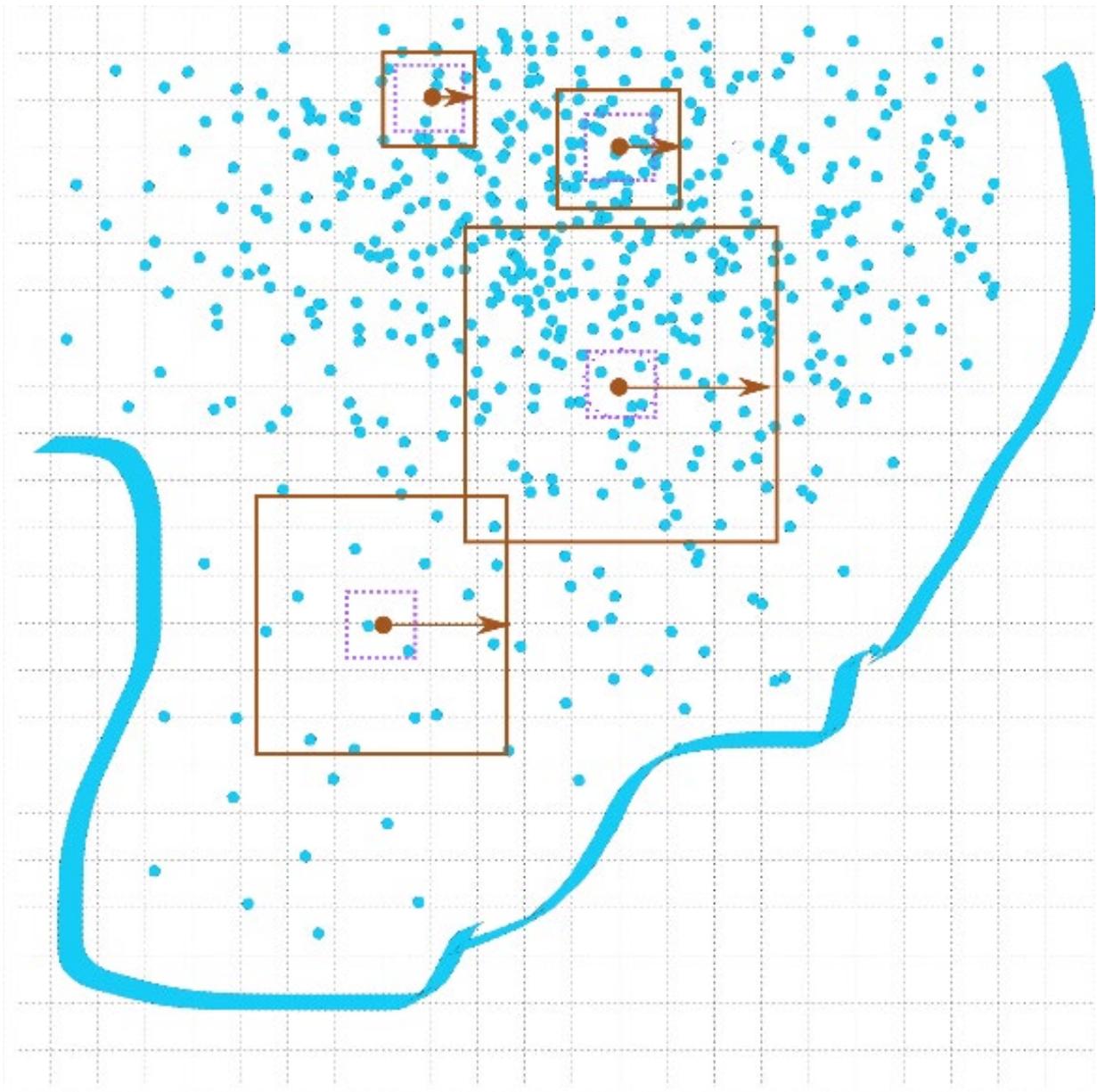


Figure 9. Répartition spatiale des données (points bleus) superposée à une grille statistique. Les m nœuds représentent les centres de chaque sous-domaine local S_m qui s'étend à une distance (flèches orientées vers la droite; carrés en ligne pleine) qui varie selon la variabilité spatiale sous-jacente des données, définie comme étant la distance à laquelle l'autocorrélation spatiale chute à une petite valeur (p. ex. $\rho_s = 0.1$). Les données situées à l'intérieur de cette distance et les paramètres obtenus à partir de l'analyse locale sont, sous l'hypothèse d'une stationnarité de second ordre, utilisés pour compléter le modèle local des processus spatiaux ou spatiotemporels puis prédire/interpoler jusqu'à une certaine fraction de la distance entre les nœuds statistiques de la grille (carré en ligne pointillée). Chaque nœud statistique est visité pour la modélisation spatiotemporelle. Toutes les prédictions qui se chevauchent sont moyennées localement (pondérées par le nombre de prédictions et la variance des prédictions). À mesure que la taille de la grille diminue, le nombre de calculs augmente en réduisant la charge de calcul et les besoins en MEV; mais l'utilité du modèle diminue également en raison de la petite taille des échantillons entrant dans les analyses. Le choix judicieux de la densité de la grille statistique ainsi que le nombre maximal et minimal de points de données et les limites supérieures et inférieures des limites spatiales doivent être équilibrés.

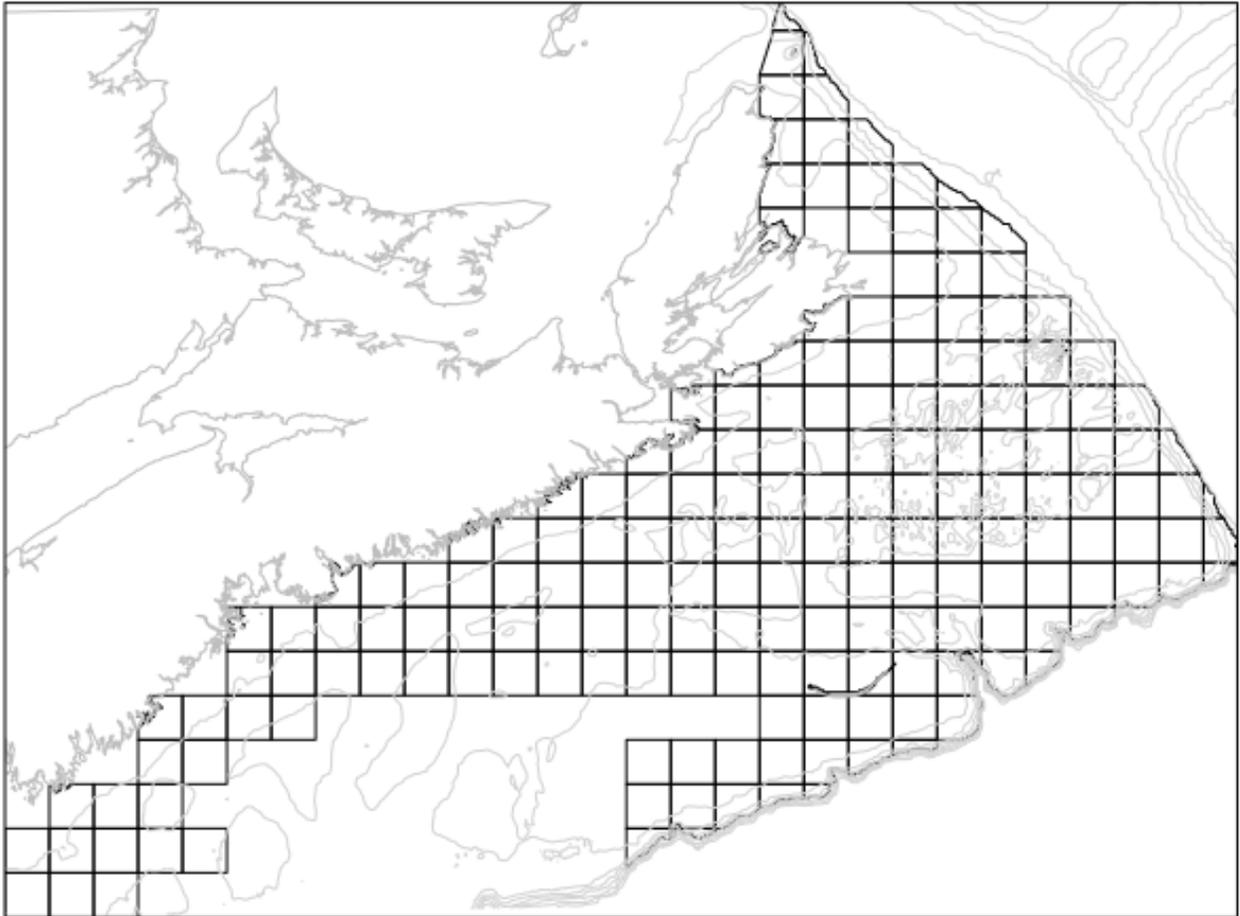


Figure 10. Unités spatiales dans la structure en treillis (25 km x 25 km) utilisée pour l'évaluation du crabe des neiges.

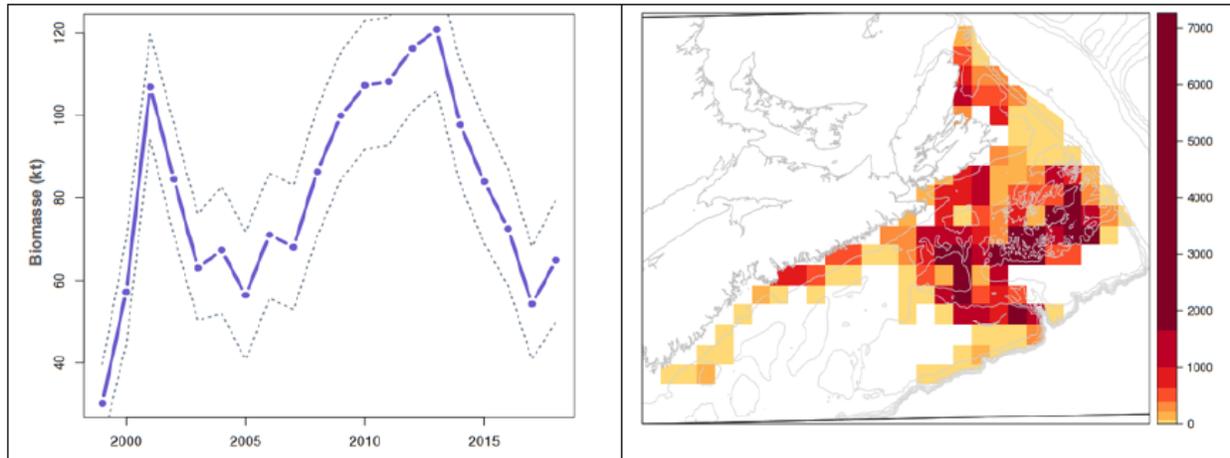


Figure 11. Variations de la biomasse (kt) du modèle « factoriel croisé gaussien (biomasse) » au fil des ans en supposant un modèle non spatial et non temporel (à gauche) avec un IC à 95 % tiré de simulations marginales. À noter que les estimations de l'abondance pour 1999, 2000, 2014 et 2017 sont probablement sous-estimées du fait que les relevés incomplets ne sont pas pris en compte dans le modèle. Estimations connexes de la biomasse dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

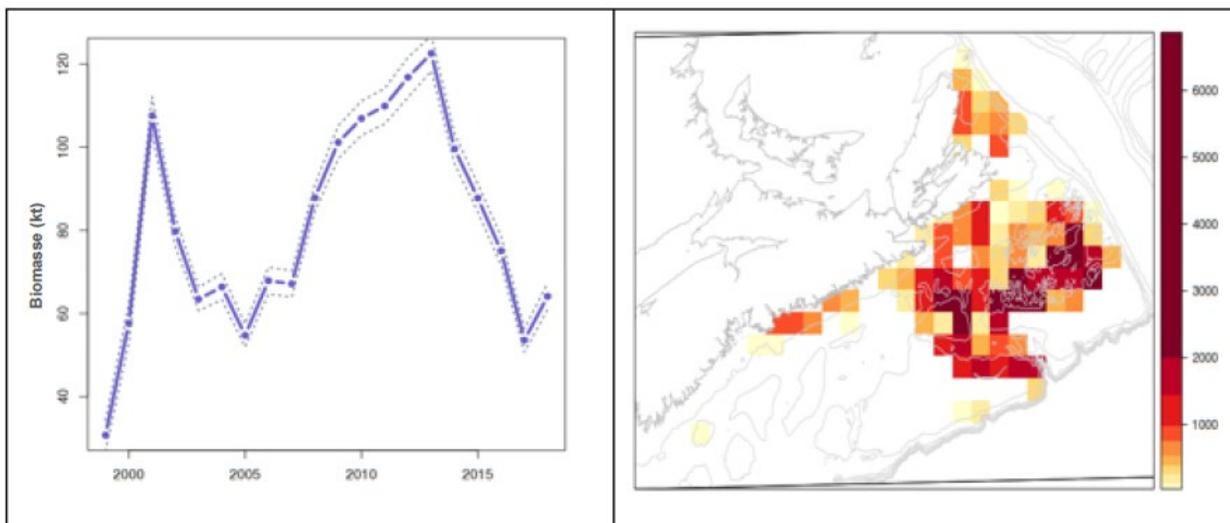


Figure 12a. Variations de la biomasse (kt) du modèle « factoriel croisé » au fil des ans en supposant un modèle non spatial et non temporel avec une hypothèse de répartition de Poisson (à gauche) avec un IC à 95 % tiré de simulations marginales. À noter que les estimations de l'abondance pour 1999, 2000, 2014 et 2017 sont probablement sous-estimées du fait que les relevés incomplets ne sont pas pris en compte dans le modèle. Estimations connexes de la biomasse dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

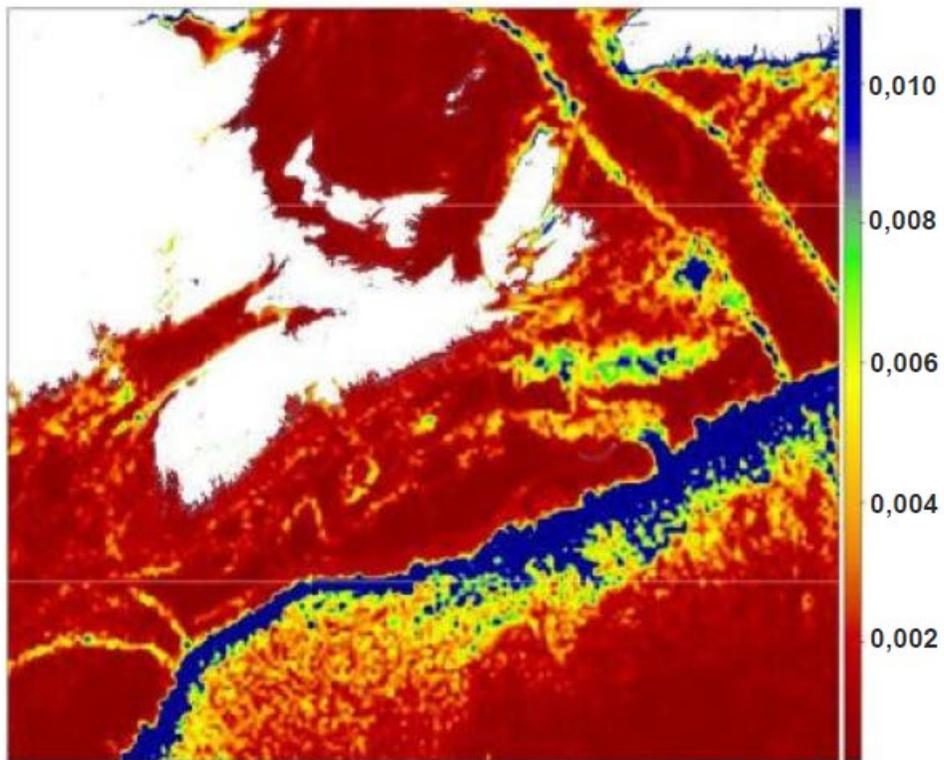
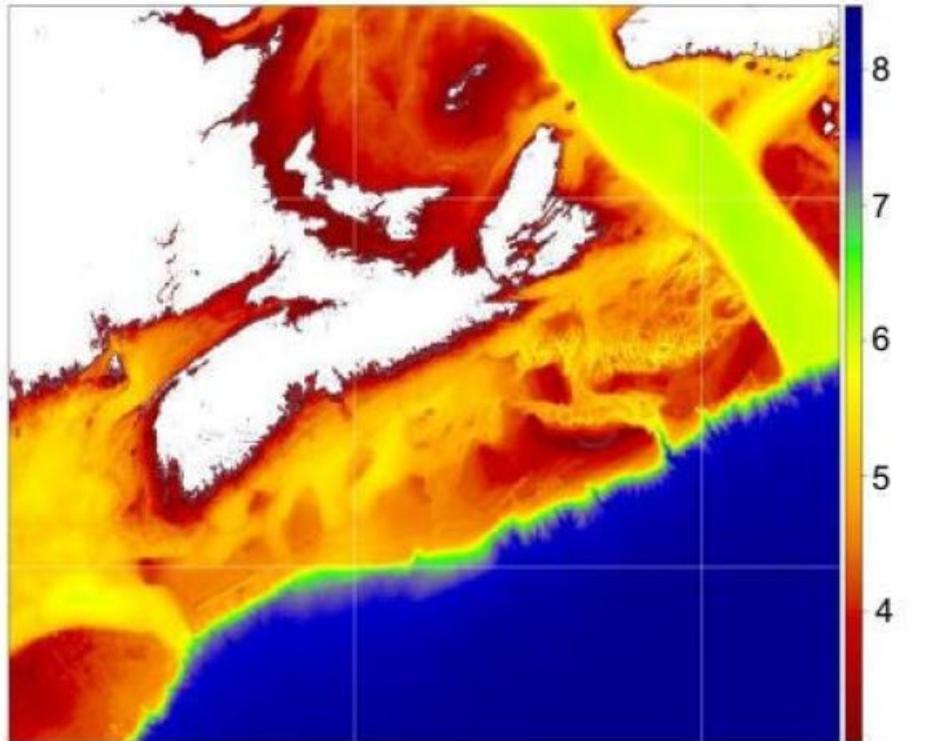


Figure 12 b. Variation de la moyenne ($\log(\text{profondeur}; m)$) et de l'écart type total ($\log(\text{profondeur})$) de la profondeur à partir d'une analyse locale (sous-domaine, c.-à-d. *stmv*). Notons que la moyenne et la variance ne sont pas uniformes dans l'espace.

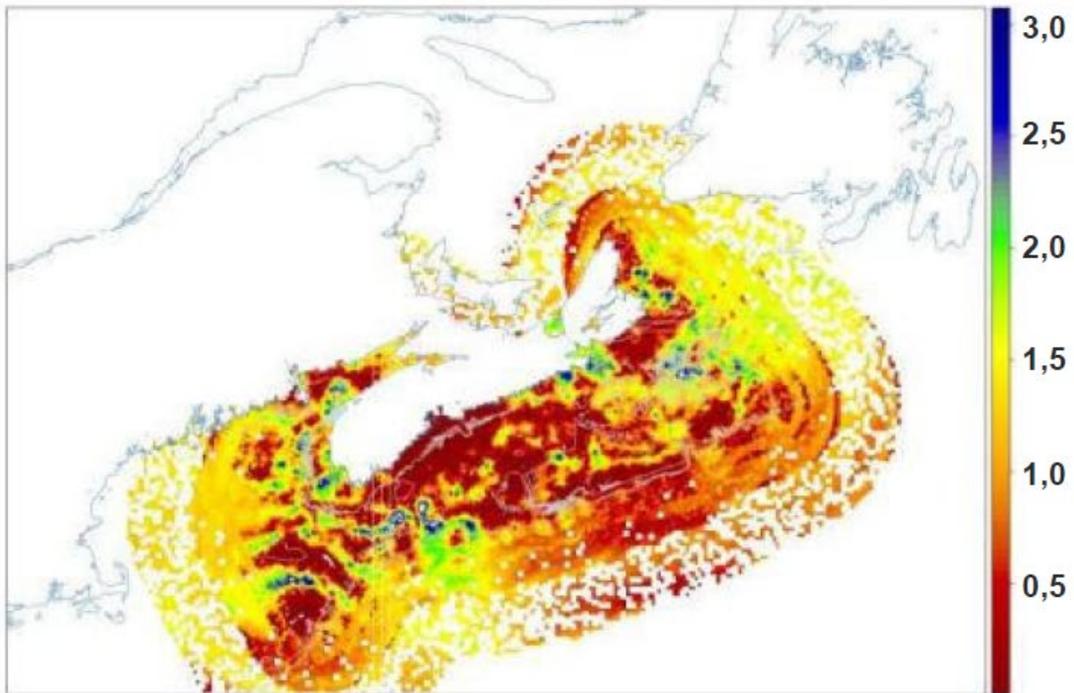
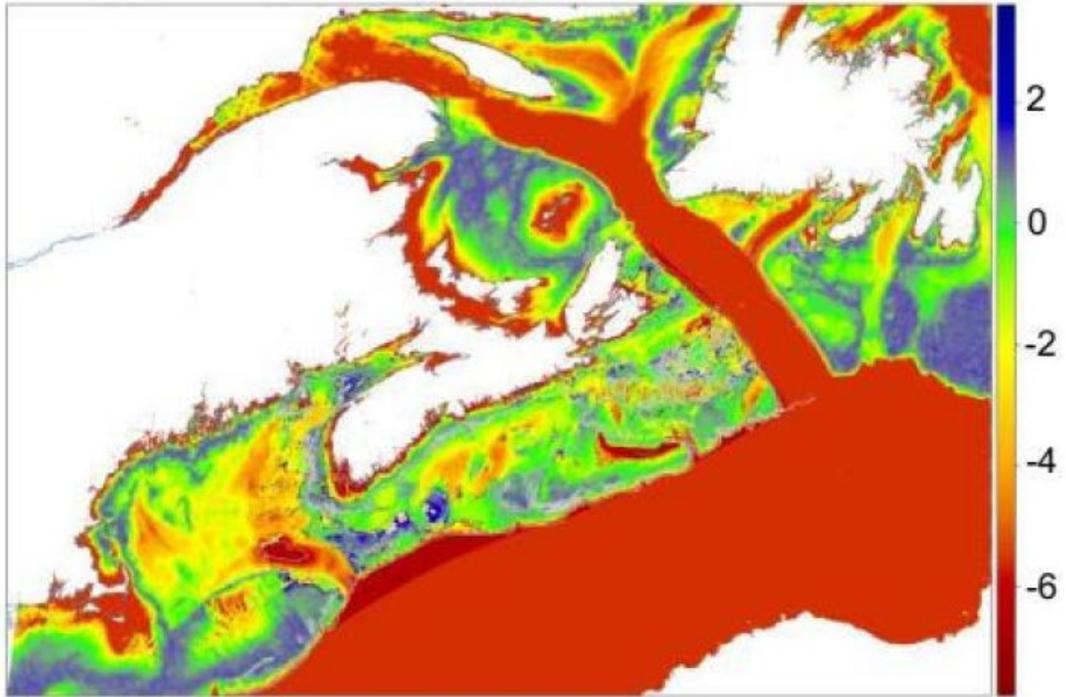


Figure 13. Variation de la moyenne et de l'écart type total de la taille des grains du substrat ($\log(\text{taille des grains du substrat, mm})$) à partir d'une analyse locale (sous-domaine, c.-à-d. *stmv*). Notons que la moyenne et la variance ne sont pas uniformes dans l'espace. Les données sont limitées à une zone plus petite comme on le voit dans *sd* (taille des grains du substrat) et au-delà de cette zone, elles sont extrapolées à partir du modèle de la moyenne qui est déterminé par la profondeur.

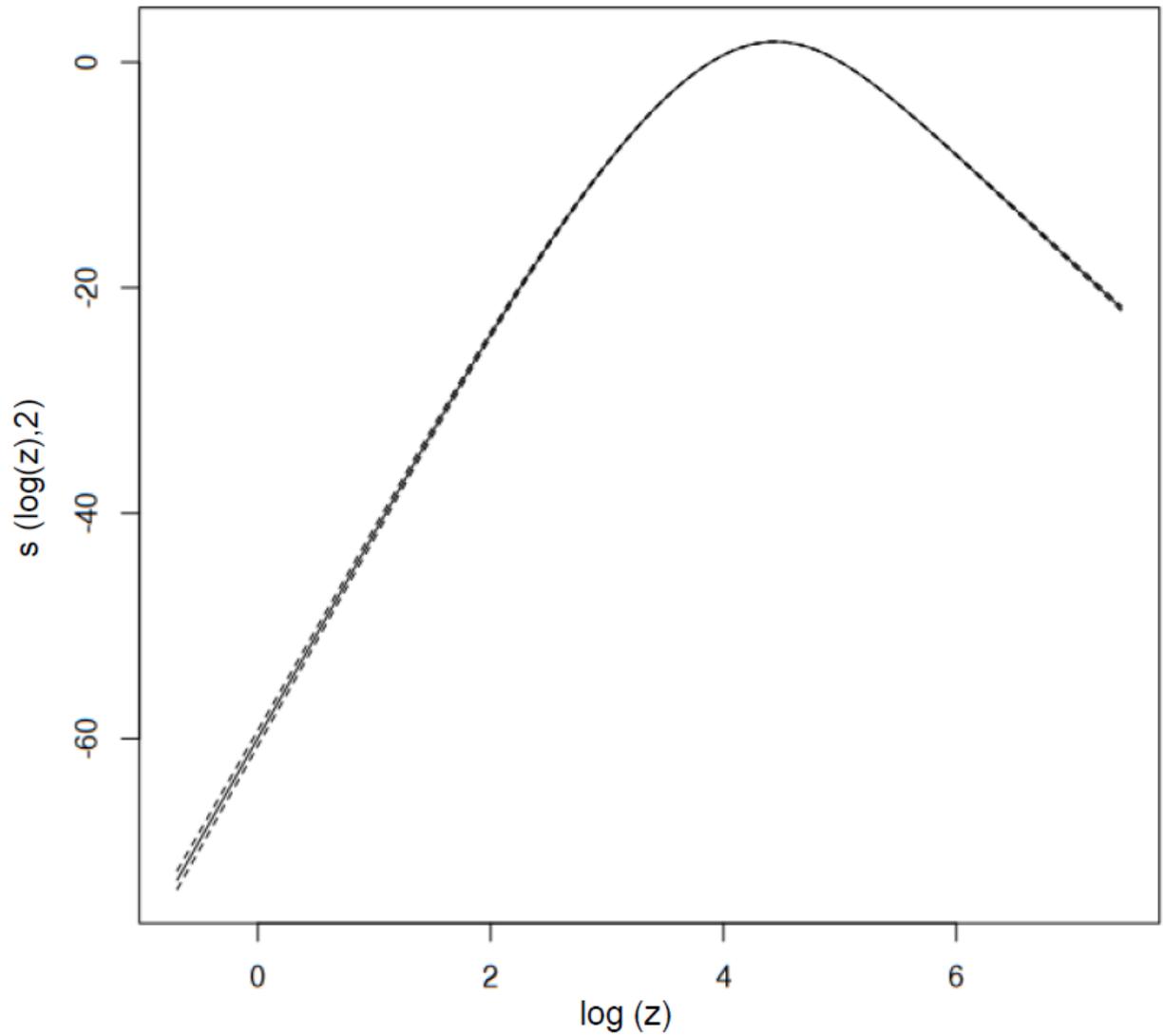


Figure 14. La taille des grains du substrat est fortement associée à la profondeur. La forme de la relation est modale avec une taille des grains de près de $\exp(4,5) = 90$ m.

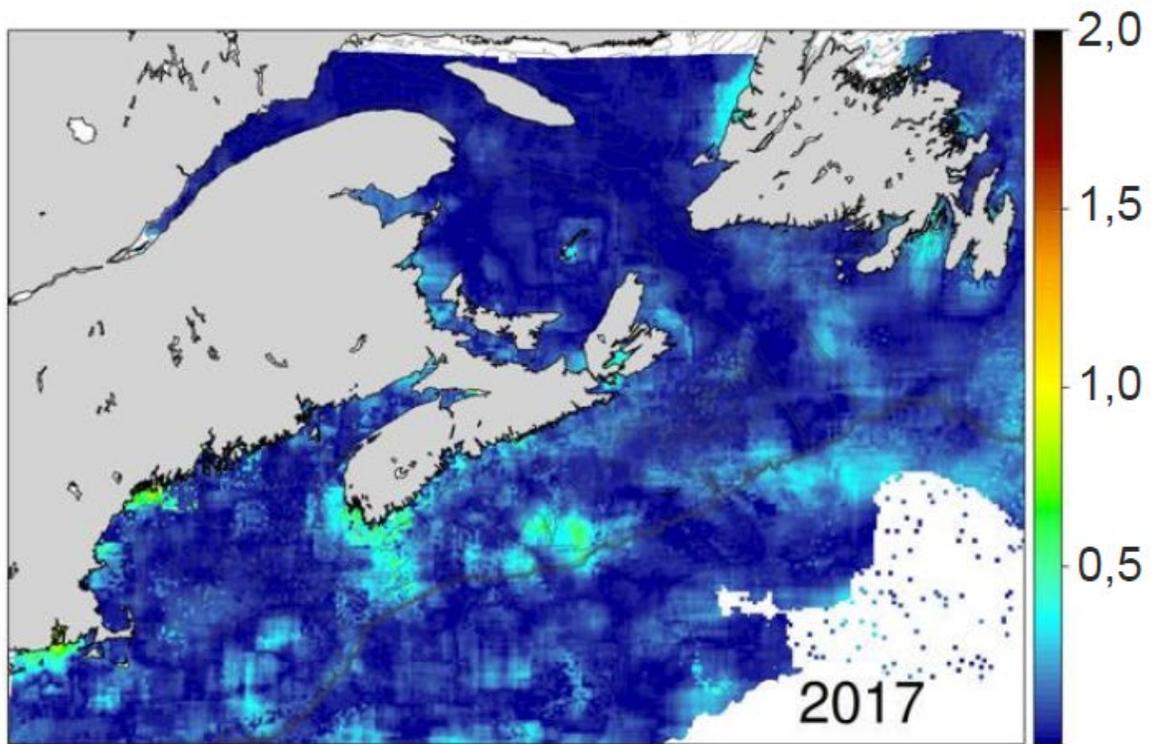
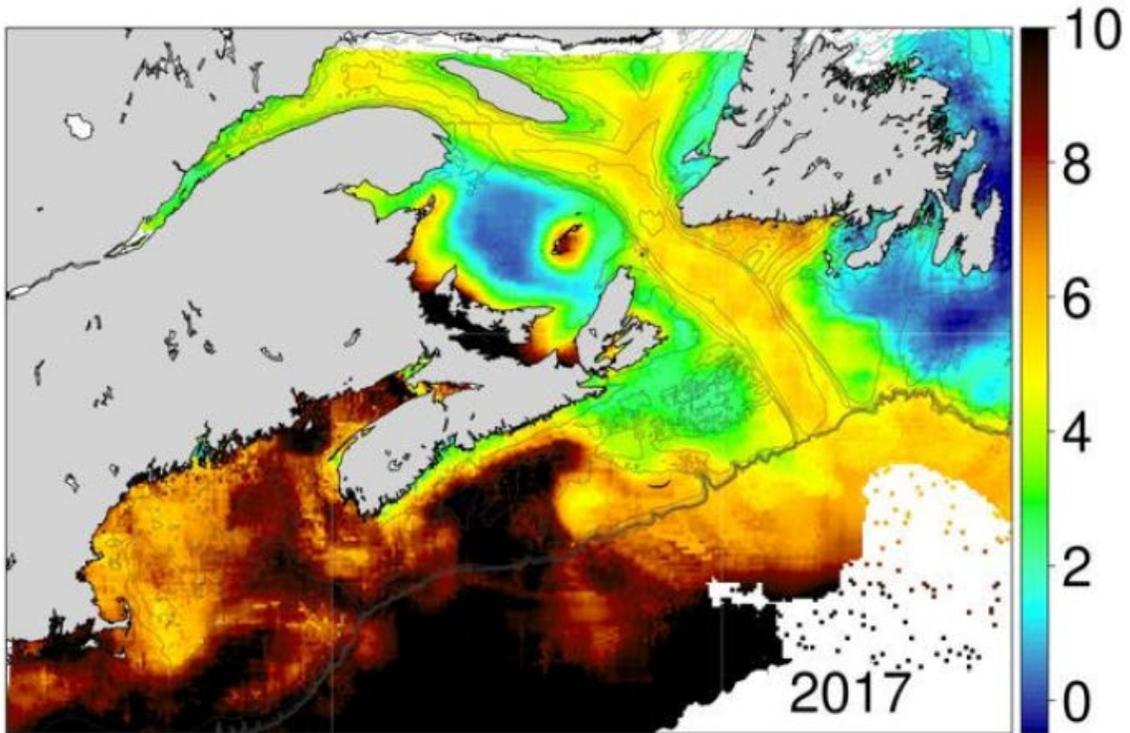


Figure 15. Variation de la moyenne et de l'écart type total de la température au fond ($^{\circ}C$) pour 2017, à partir d'une analyse locale (sous-domaine, c.-à-d. stm). Notons que la moyenne et la variance ne sont pas uniformes dans l'espace.

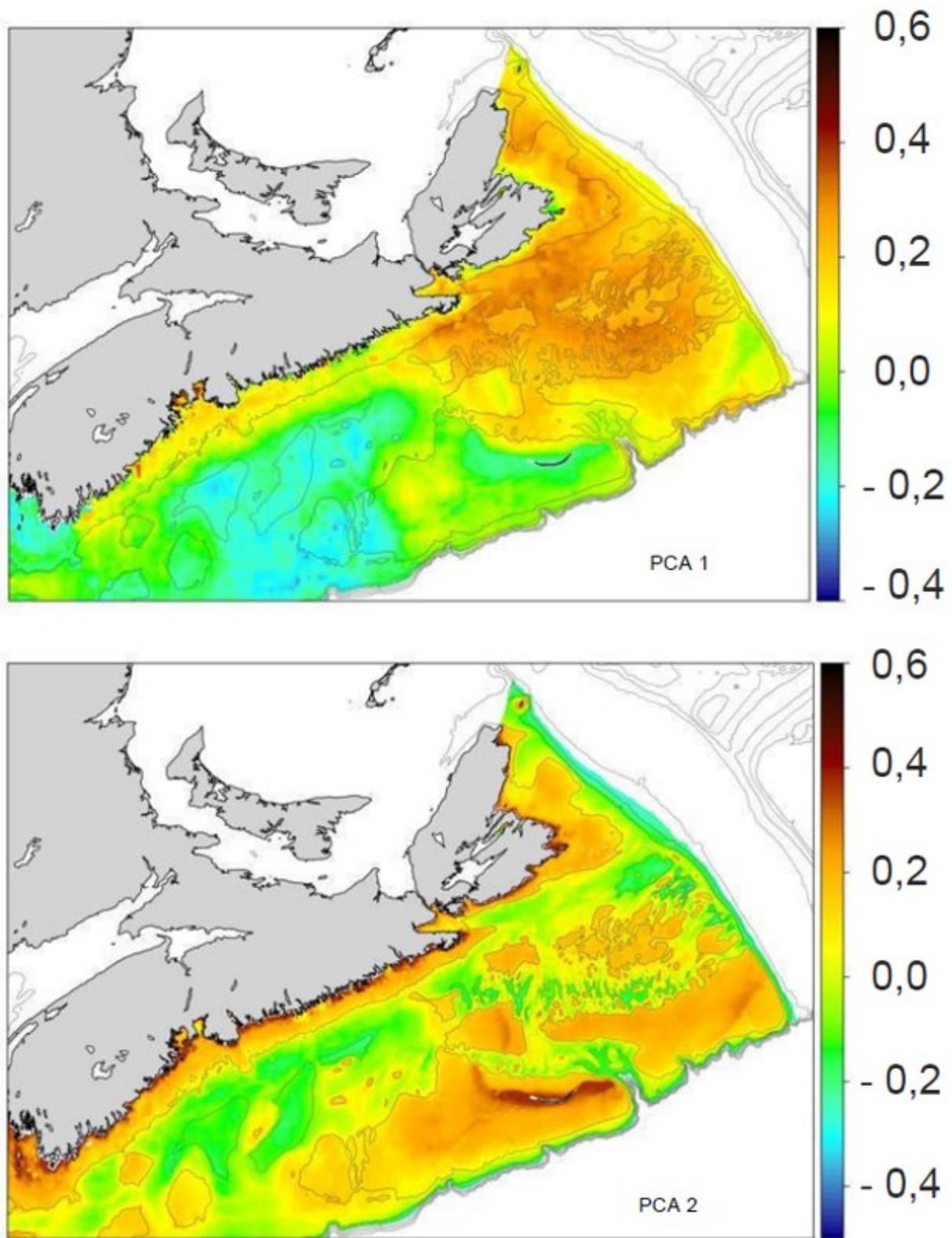


Figure 16. Variation des scores moyens des axes 1 et 2 de la composition de l'espèce pour 2017, à partir d'une analyse locale (sous-domaine, c.-à-d. *stmv*). Notons que la moyenne (et la variabilité connexe) n'est pas uniforme dans l'espace.

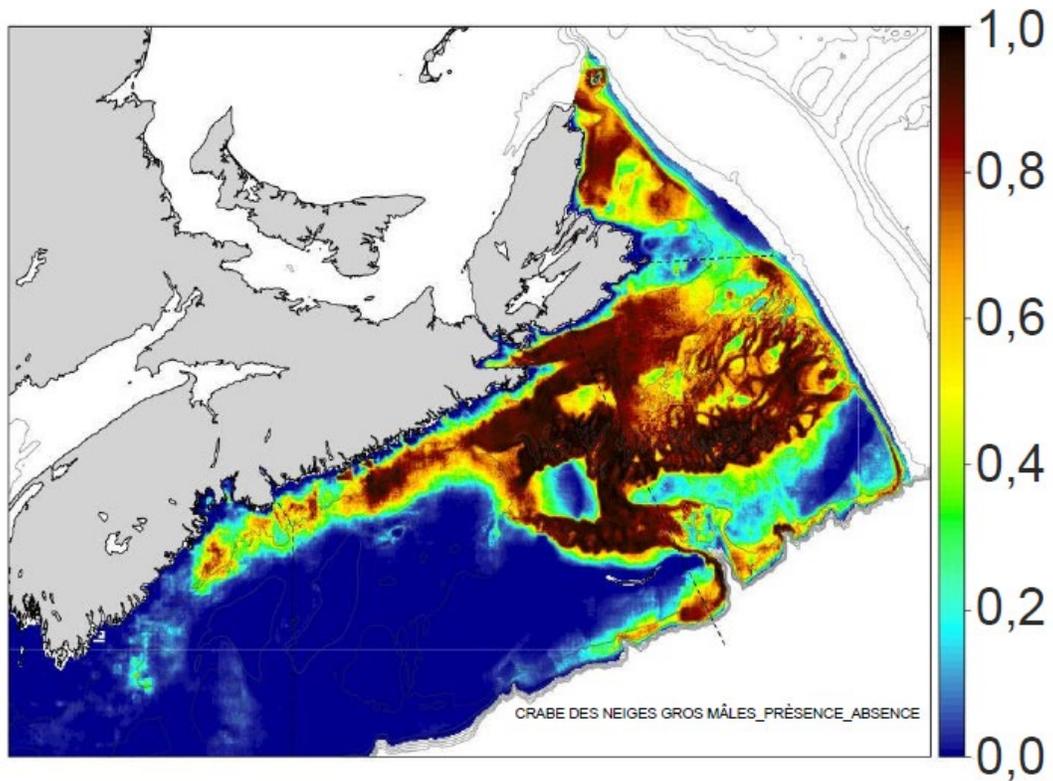


Figure 17. Probabilité d'observer le crabe des neiges pour 2017, à partir d'une analyse locale (sous-domaine, c.-à-d. stmv).

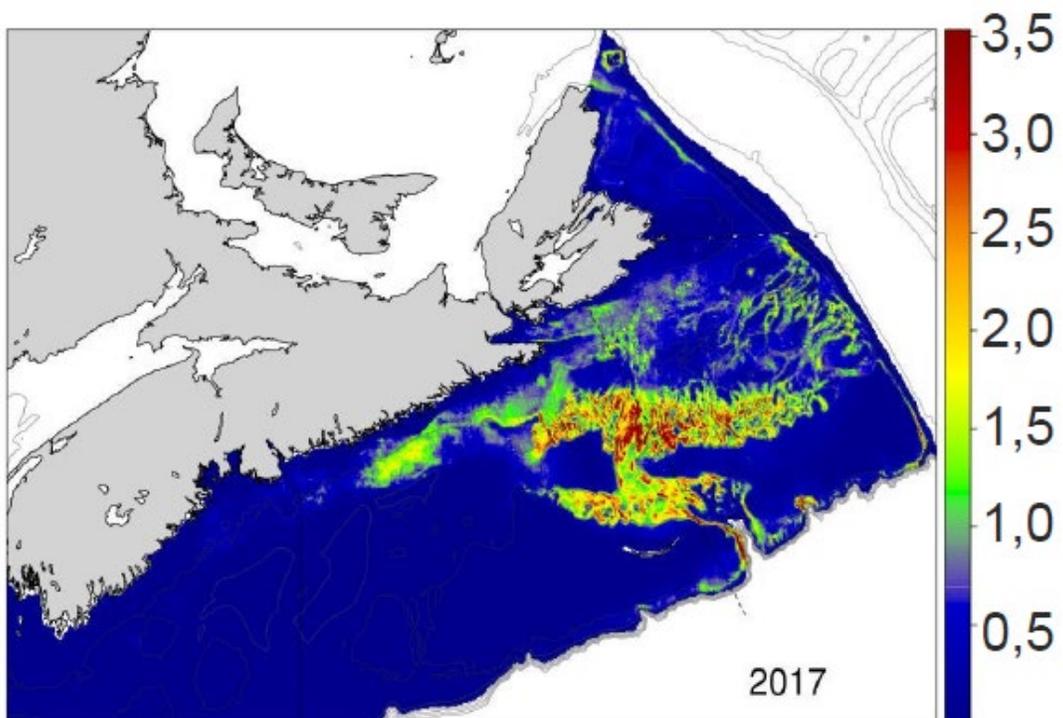


Figure 18. Densité de biomasse du crabe des neiges pour 2017, à partir d'une analyse locale (sous-domaine, c.-à-d. stmv).

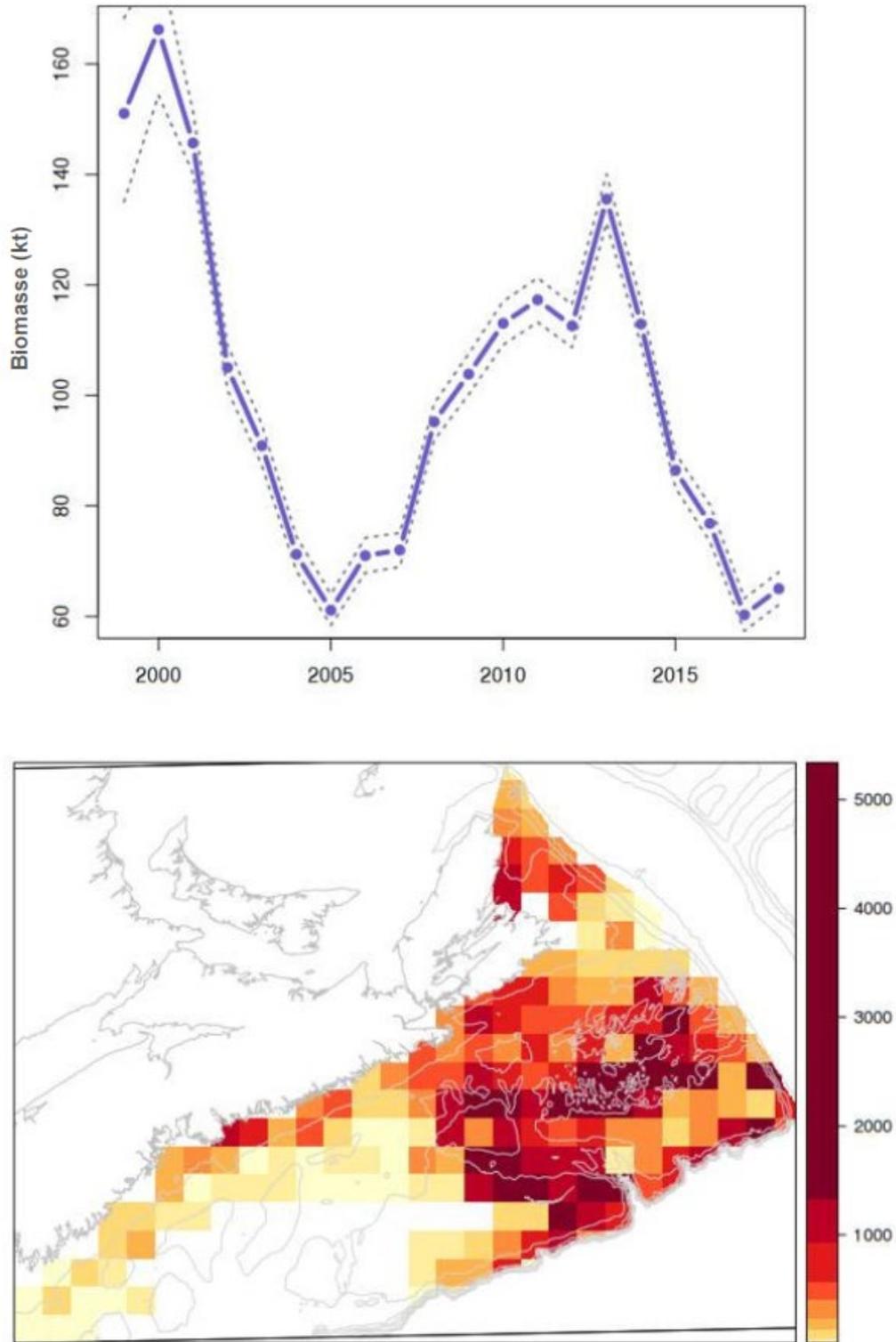


Figure 19. Modèle simple à effets mixtes. Variations de la biomasse (kt) au fil des années en supposant un effet aléatoire CAR unique, et les années comme modèle à effet fixe avec hypothèse de répartition de Poisson (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes. Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

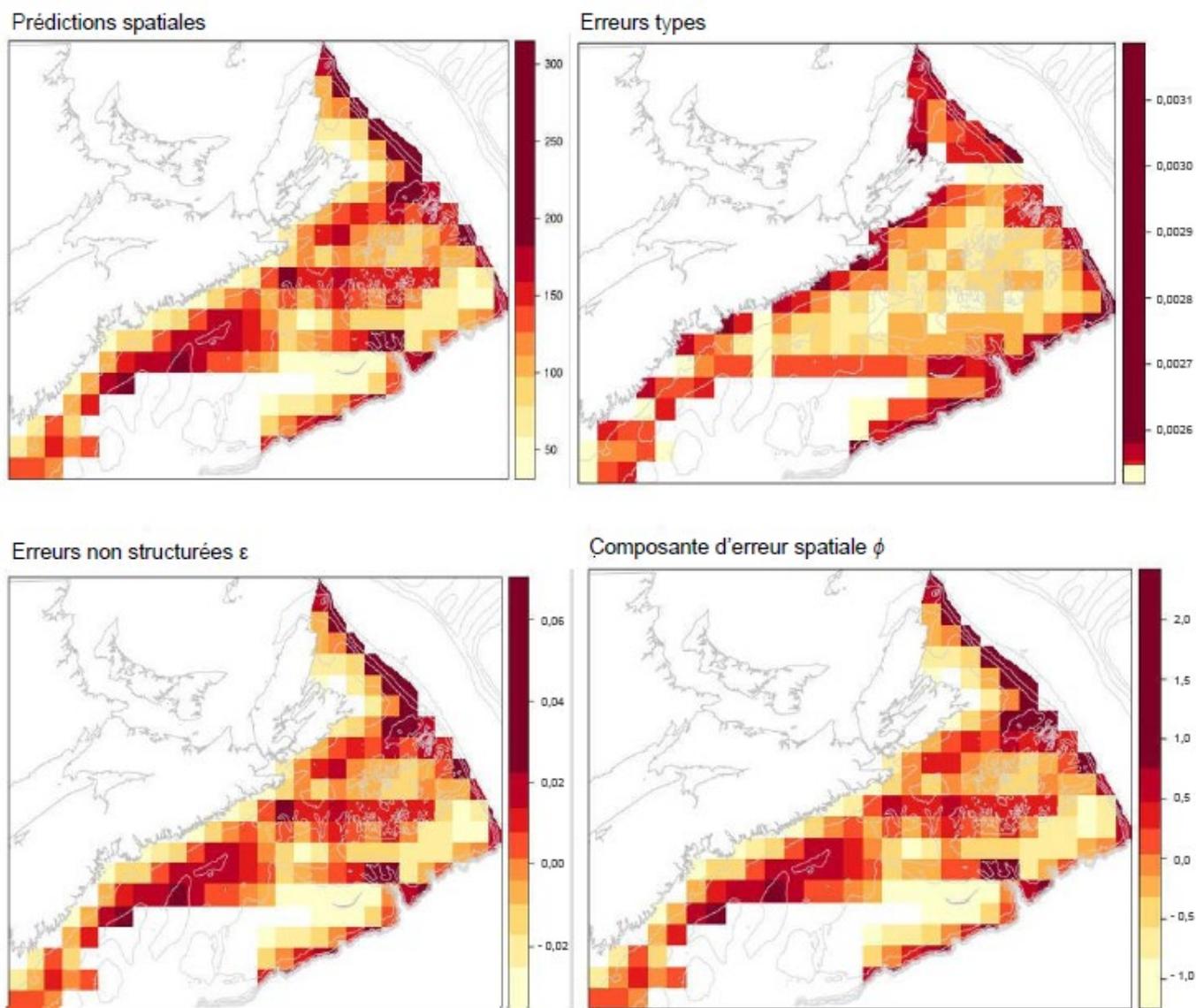


Figure 20. Bathymétrie (m) – analyse CAR : prédictions spatiales (en haut à gauche), erreurs types (en haut à droite), erreurs non structurées ϵ (en bas à gauche) et composantes d'erreur spatiale ϕ (en bas à droite).

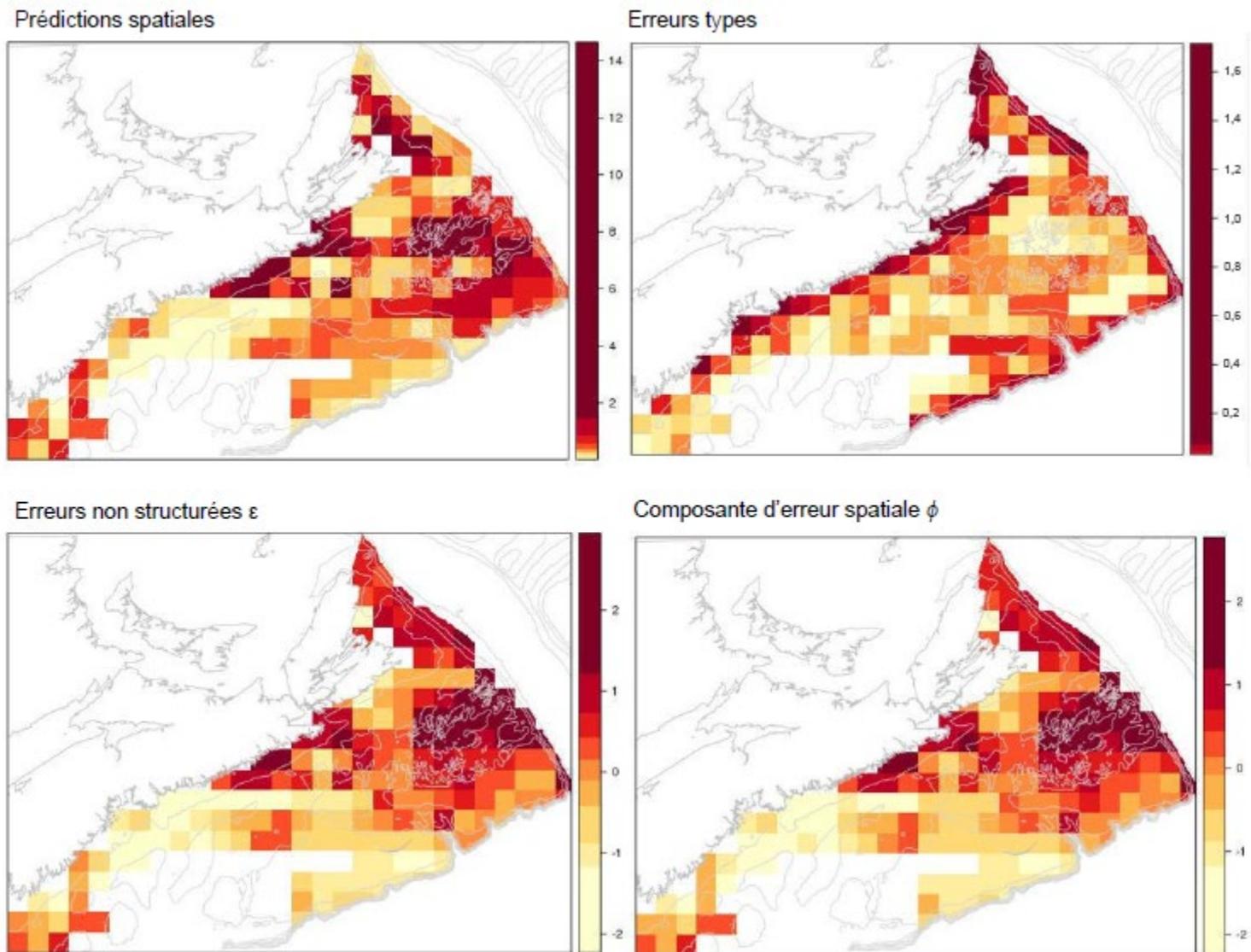


Figure 21. Taille des grains du substrat (mm) – analyse CAR : prédictions spatiales (en haut à gauche), erreurs types (en haut à droite), erreurs non structurées ϵ (en bas à gauche) et composantes d'erreur spatiale ν (en bas à droite).

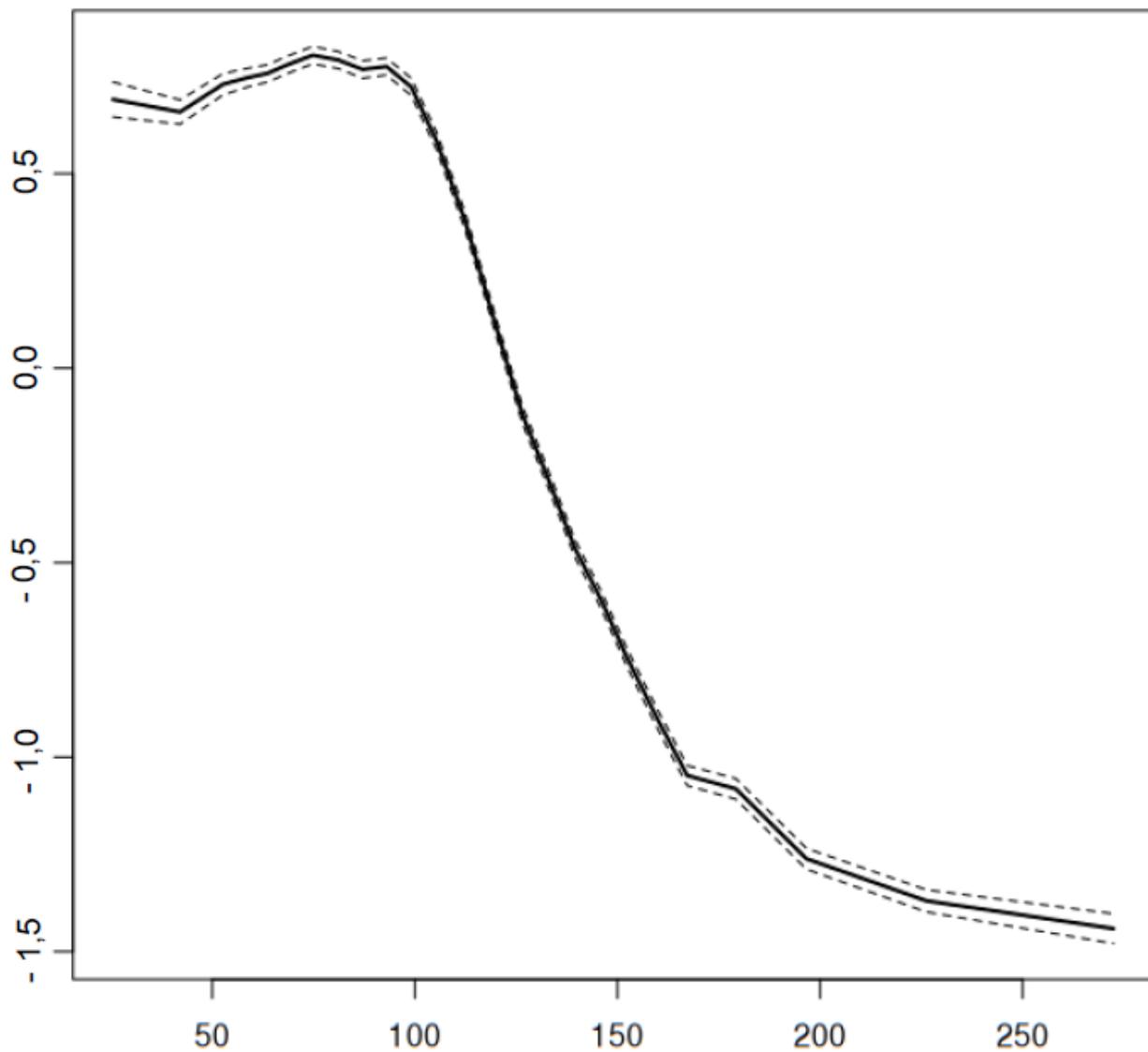


Figure 22. Rapport entre la taille des grains du substrat (log, mm) et la profondeur (m). La plus grande variation de la taille des grains du substrat est observée entre 100 m et 200 m, au-delà de cette profondeur il se stabilise.

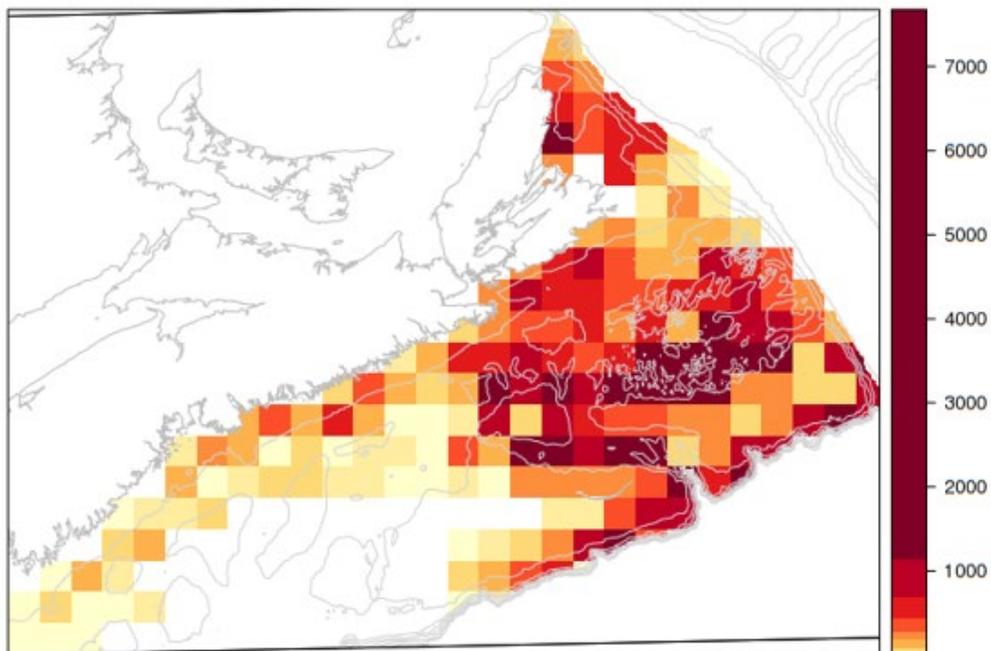
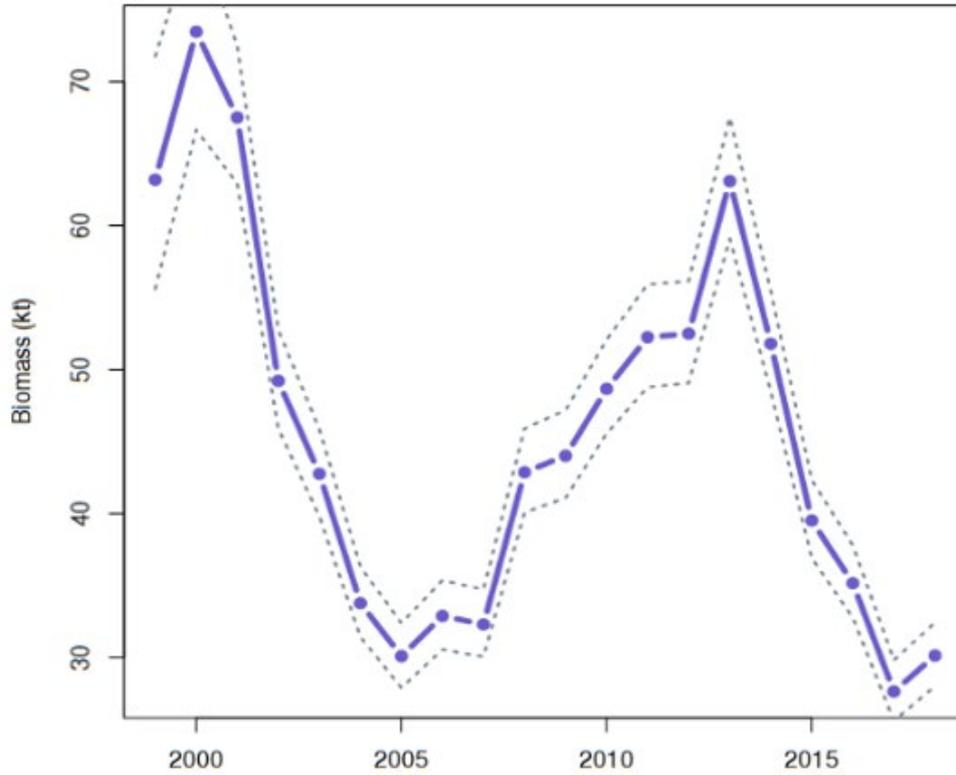


Figure 23. Résultats du modèle « statique à effets mixtes ». Variations de la biomasse (kt) fondées sur un modèle de Poisson au fil des années en supposant un effet aléatoire CAR unique et la taille des grains du substrat et la profondeur comme covariables avec une hypothèse de répartition de Poisson (en haut). Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

`inla.group(z, method = "quantile", n = 13)`

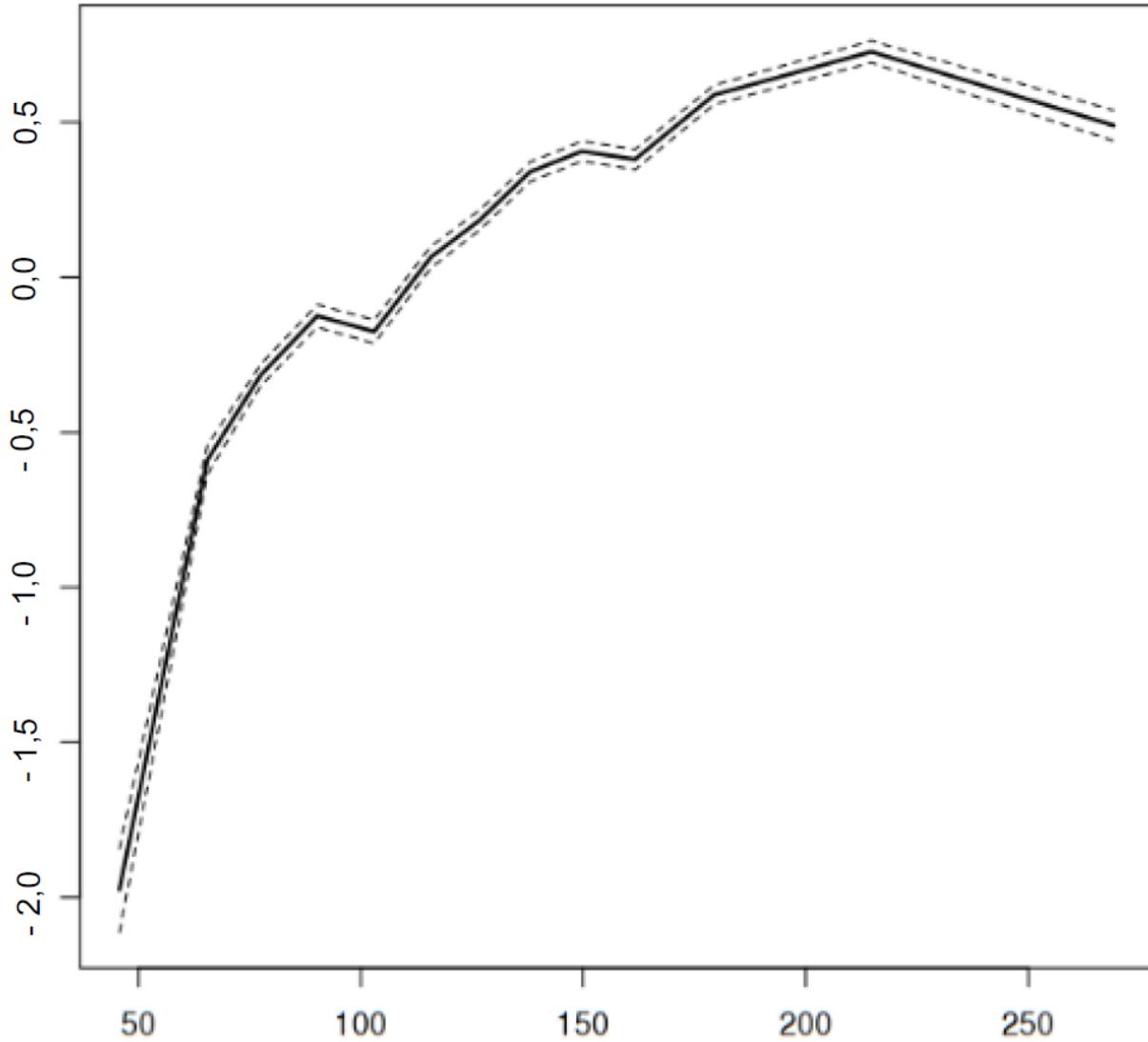


Figure 24. Densité numérique c. profondeur (m) après ajustement des erreurs fondées sur CAR (centré sur la moyenne globale : $6\,723/km^2$).

`inla.group(substrate.grainsize, method = "quantile", n = 13)`

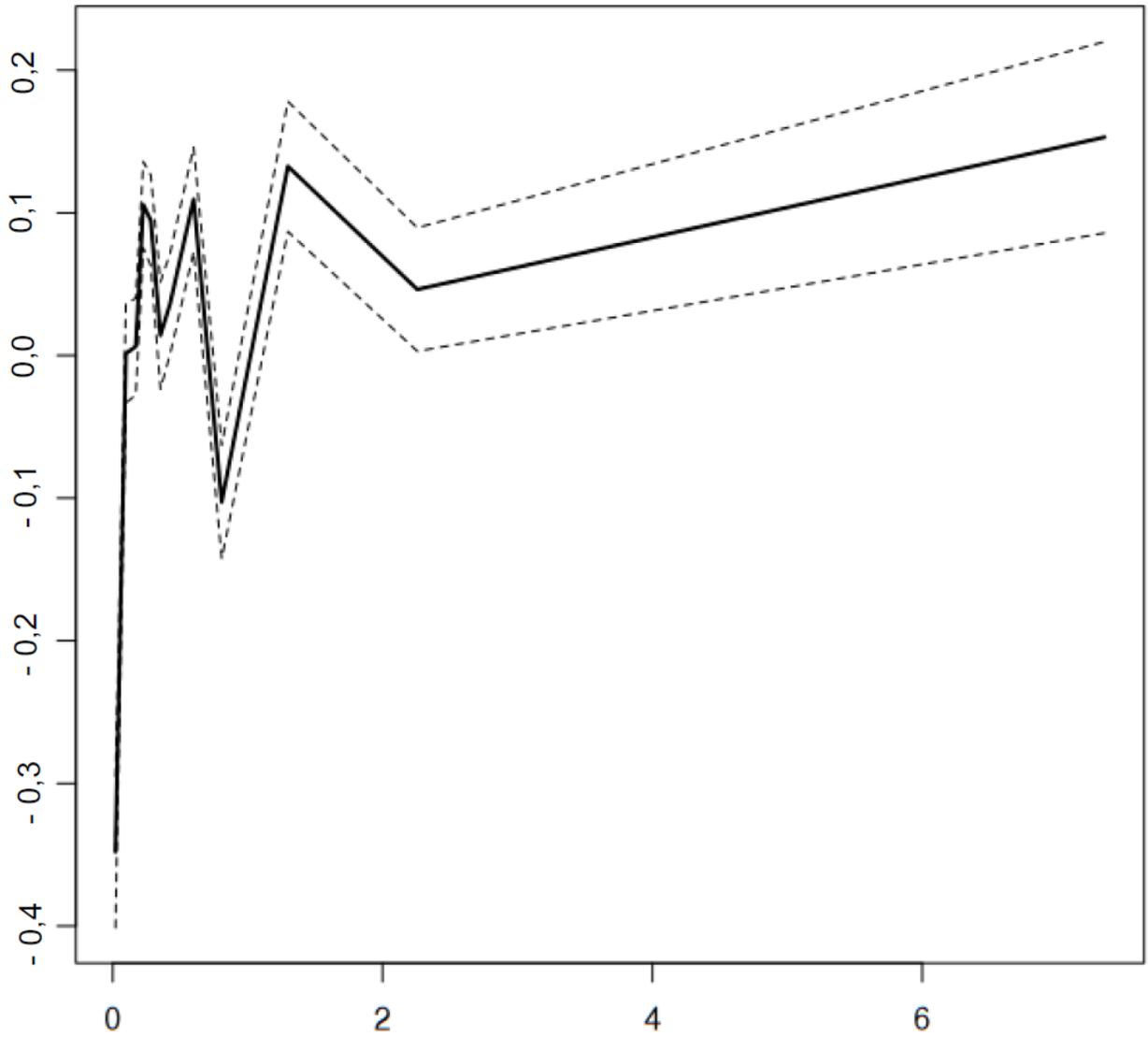


Figure 25. Densité numérique c. taille des grains du substrat (mm) après ajustement des erreurs fondées sur CAR (centré sur la moyenne globale : 6 723/km²).

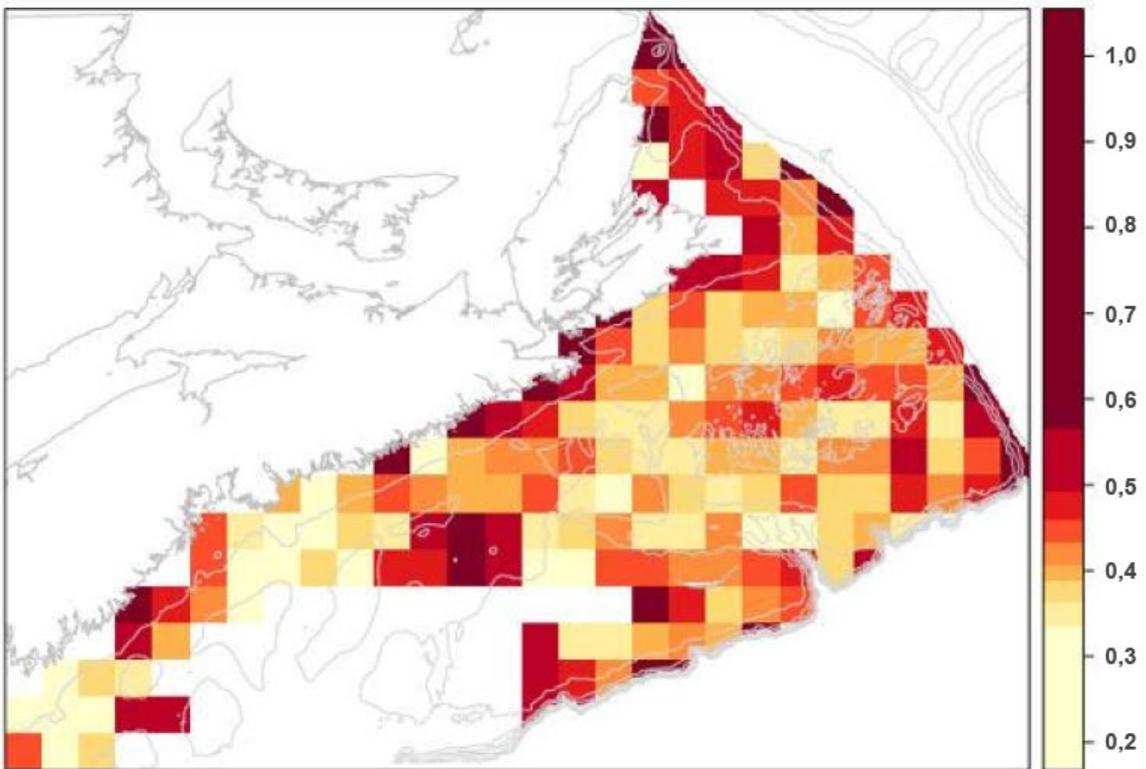
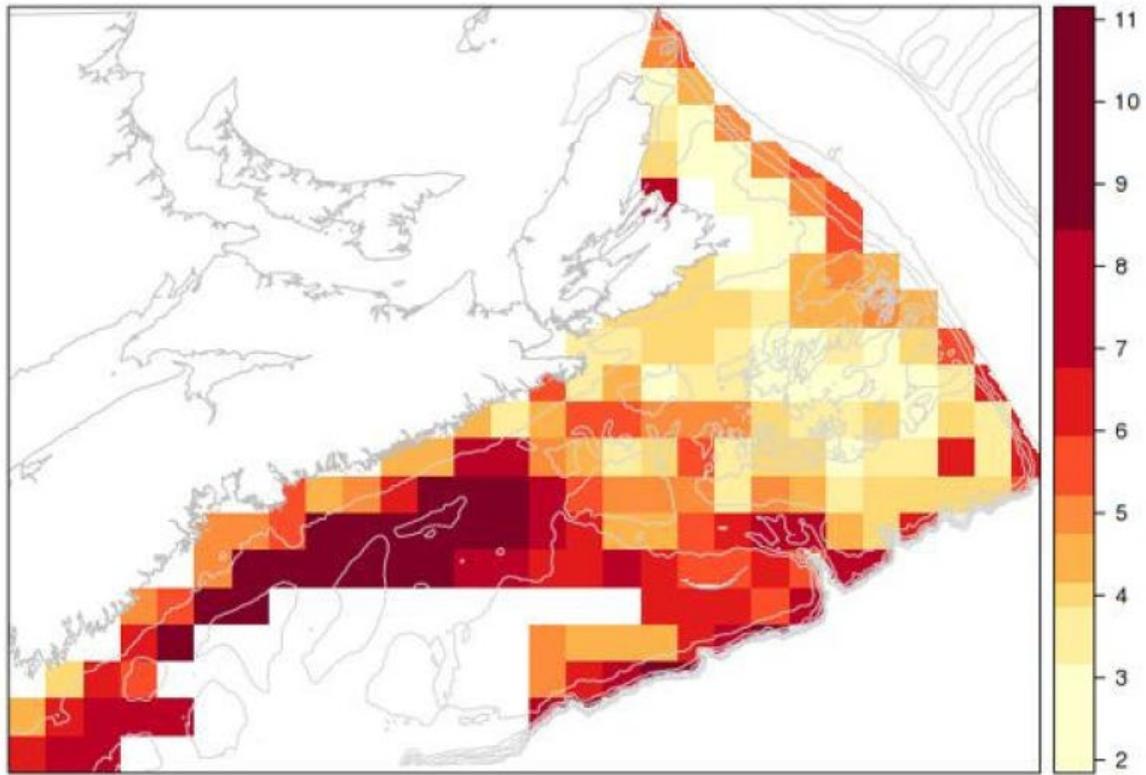


Figure 26. Température au fond ($^{\circ}C$) – analyse CAR : pour le début d'octobre 2000, prédictions spatiales (en haut) et erreurs types (en bas).

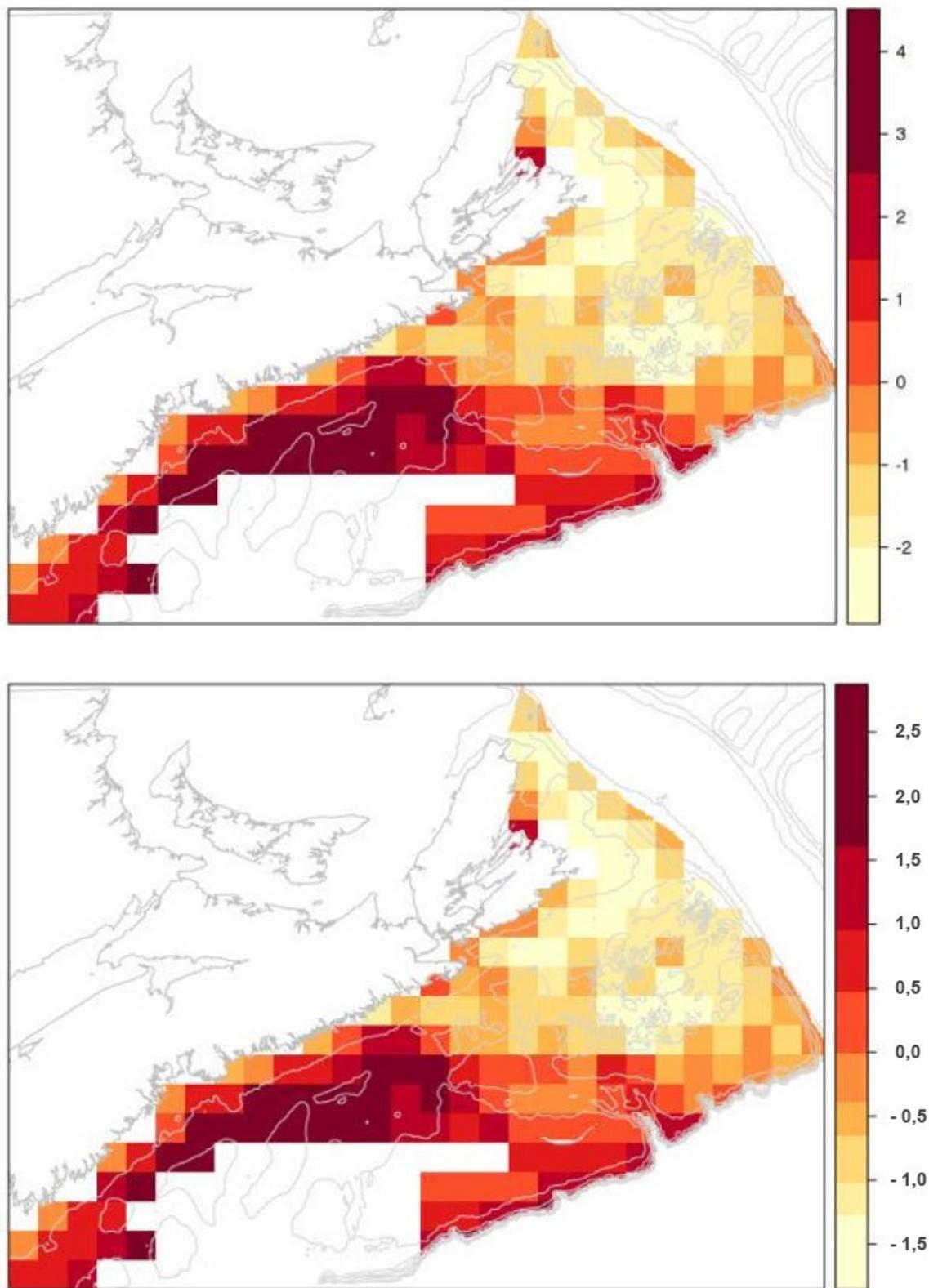


Figure 27. Température au fond ($^{\circ}C$) – analyse CAR : pour l’année 2000, erreurs non structurées ε (en haut) et composantes d’erreur spatiale ϕ (en bas). Le mode a posteriori pour ρ , la proportion d’erreur attribuable aux effets spatiaux, est de 0,983 (ICB à 95 % : 0,983; 1).

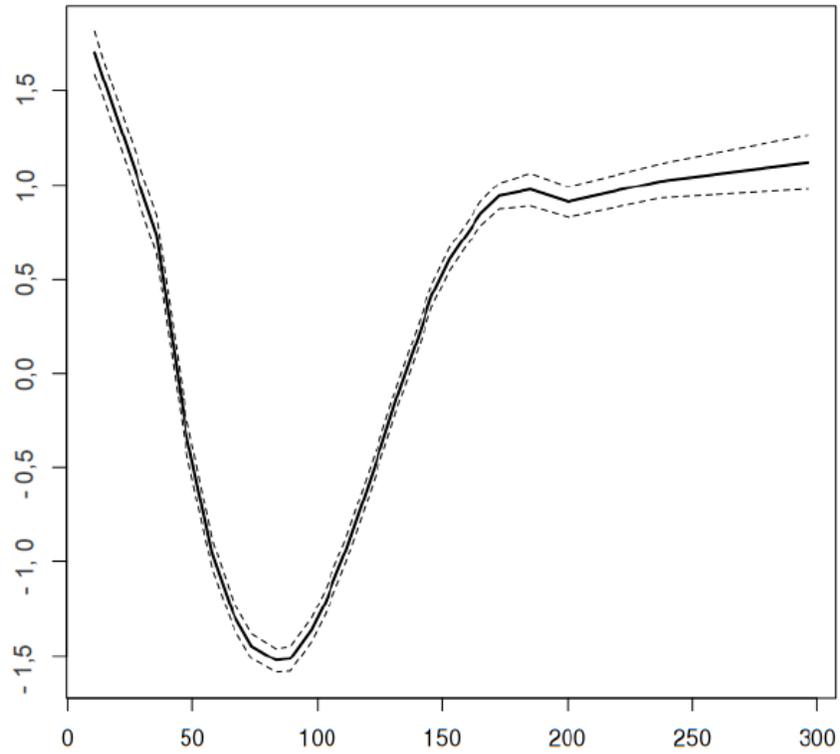


Figure 28. Température au fond (°C) c. profondeur (m) après ajustement des erreurs fondées sur CAR (centré sur la moyenne globale : 4 52°C).

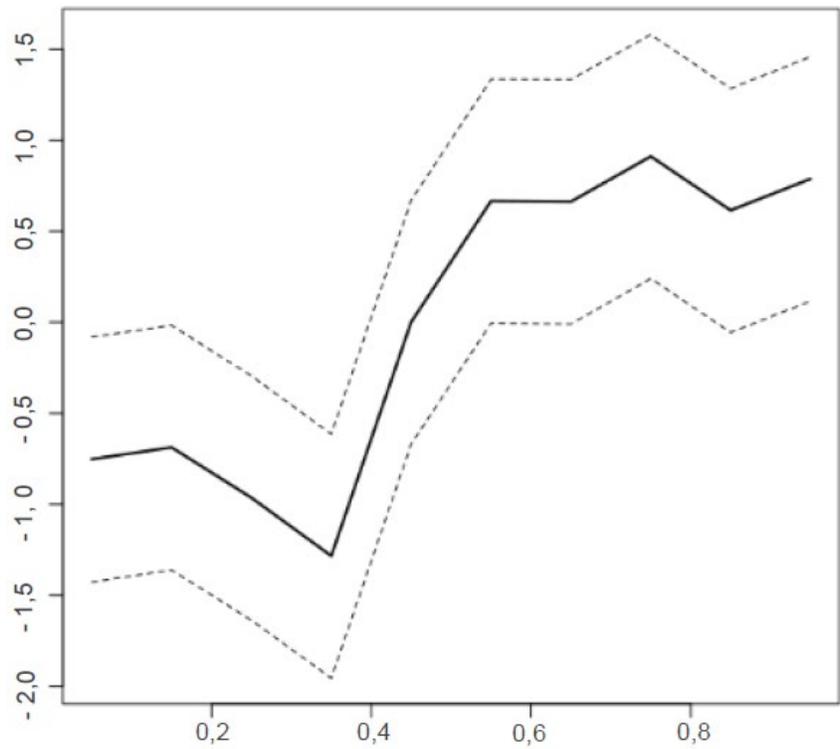


Figure 29. Variations des températures au fond – tendances saisonnières (centré sur la moyenne globale : 4 52°C).

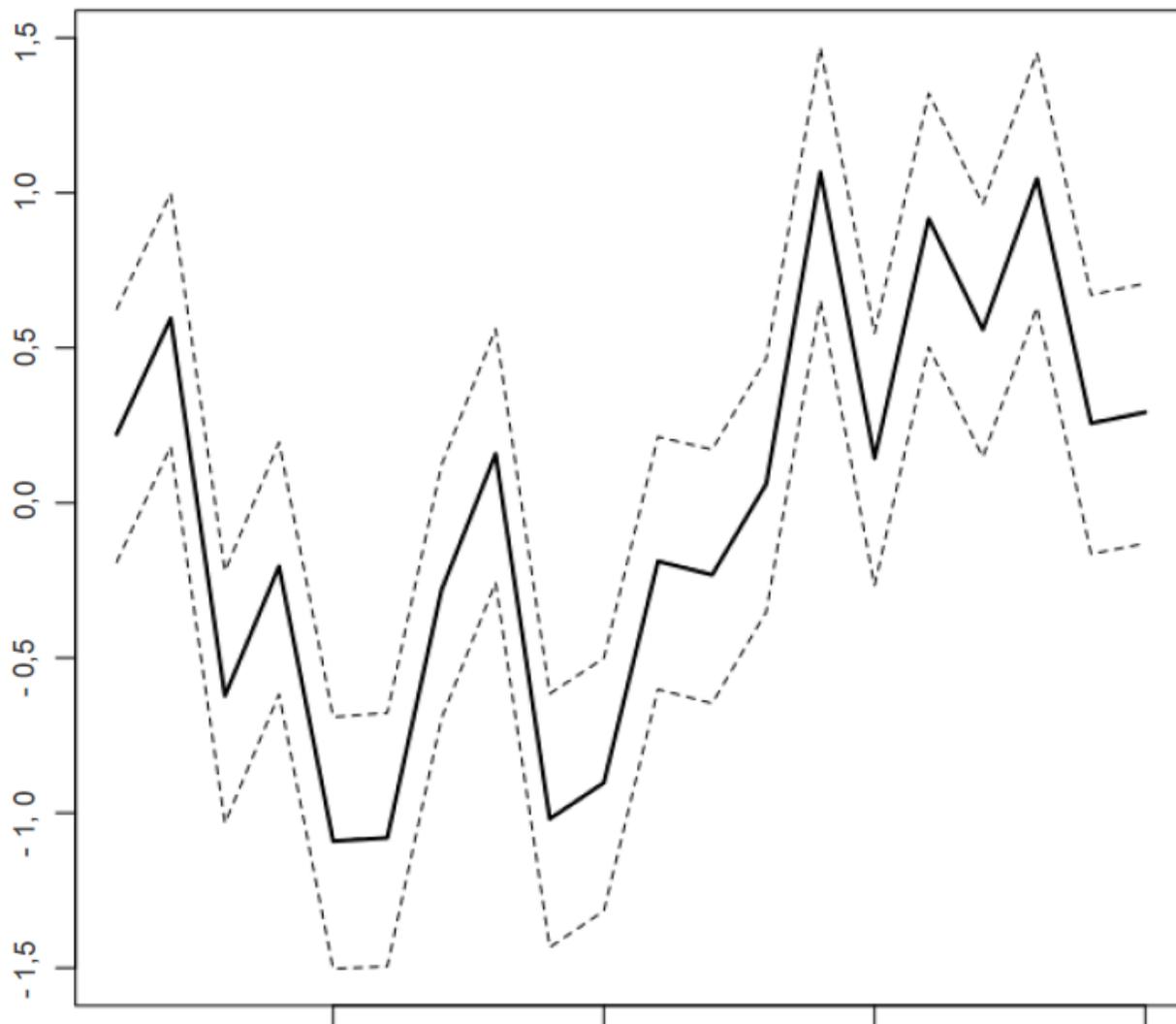


Figure 30. Variations des températures au fond – tendances annuelles de 1999 à 2018 (centré sur la moyenne globale : 4 52° C). Cela indique que les températures ont été élevées au cours des huit dernières années.

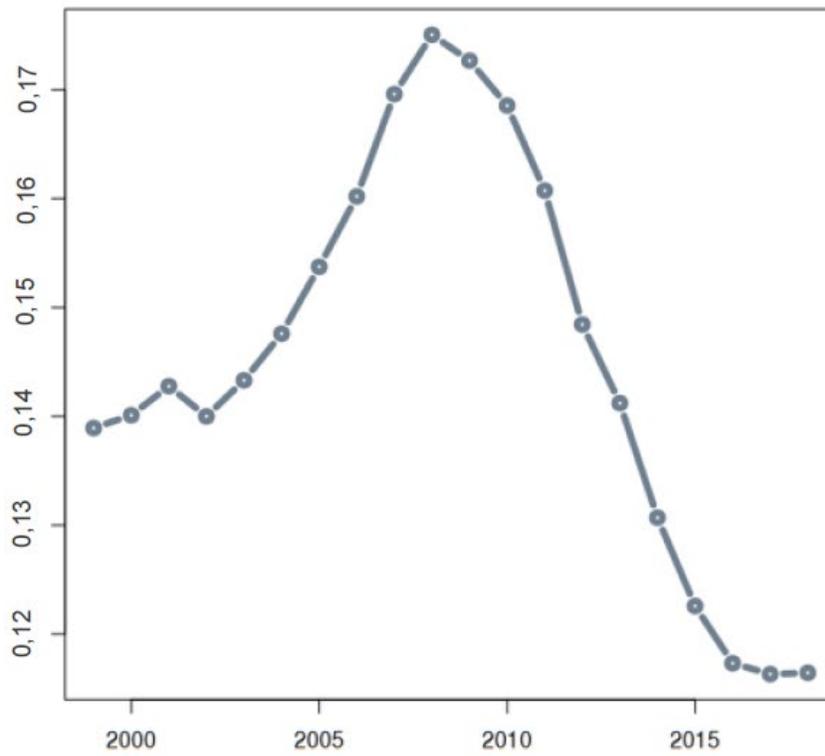
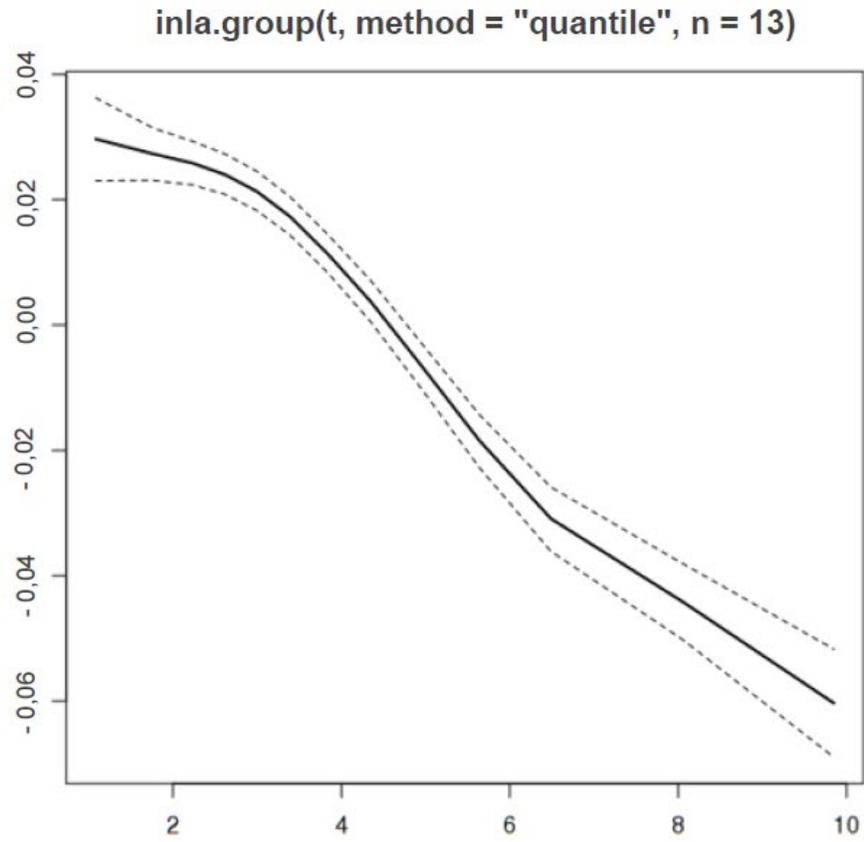


Figure 31. Axe 1 de la composition de l'espèce c. température (en haut) et temps (en bas).

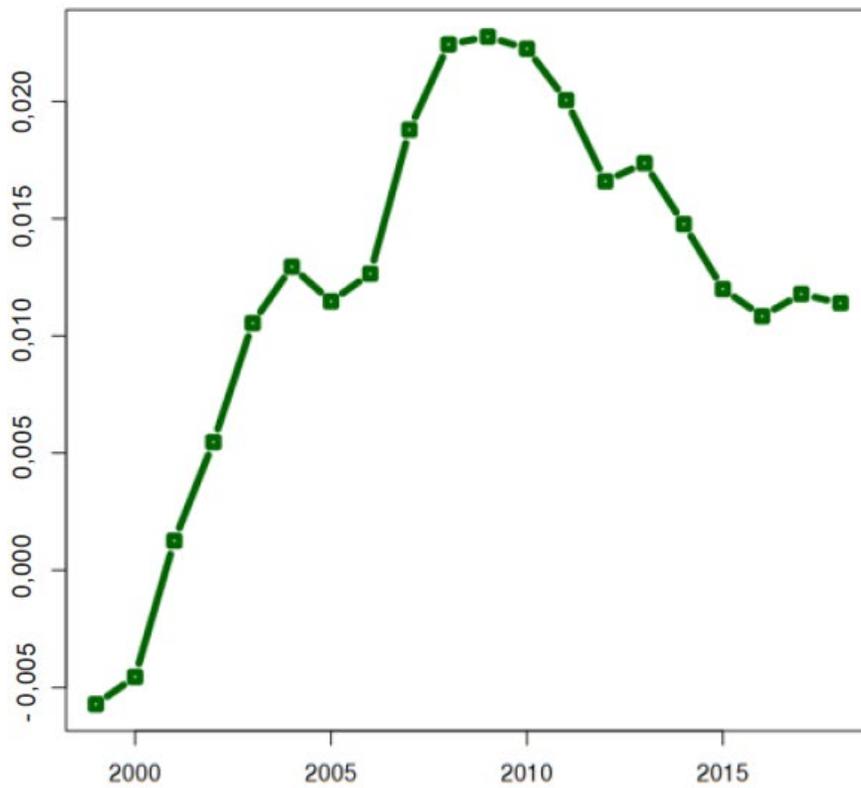
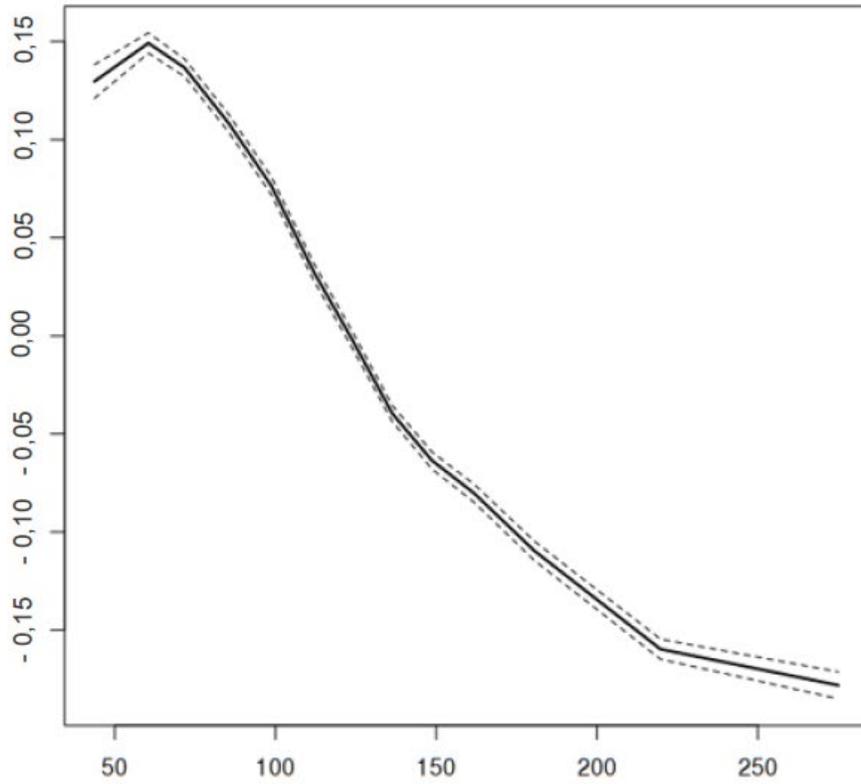


Figure 32. Axe 2 de la composition de l'espèce *c.* profondeur (en haut) et temps (en bas).

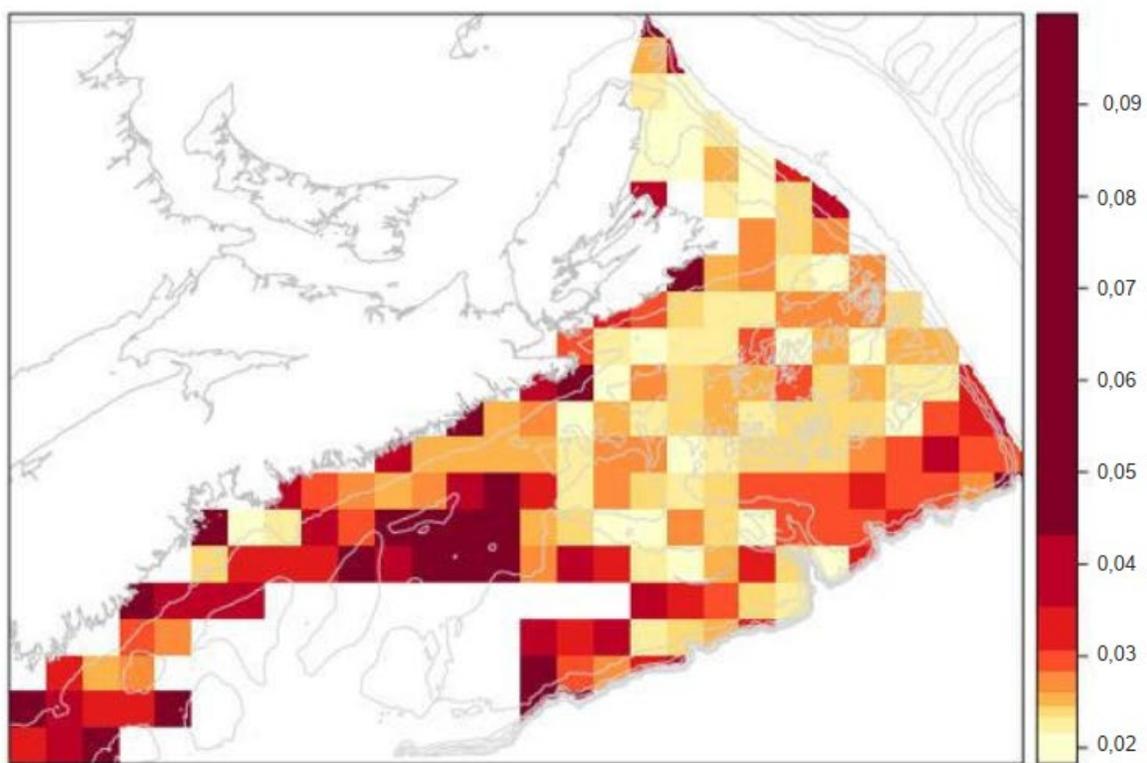
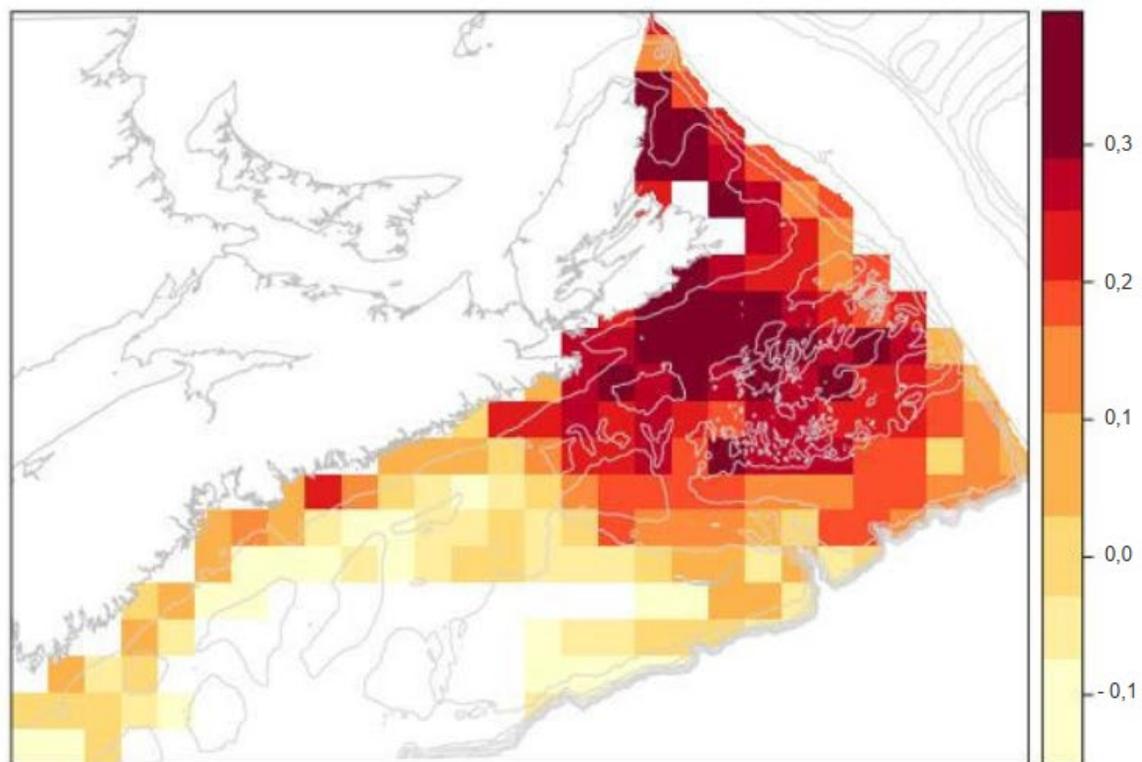


Figure 33. Axe 1 de la composition de l'espèce pour septembre 2017; moyenne (en haut) et écart type (en bas).

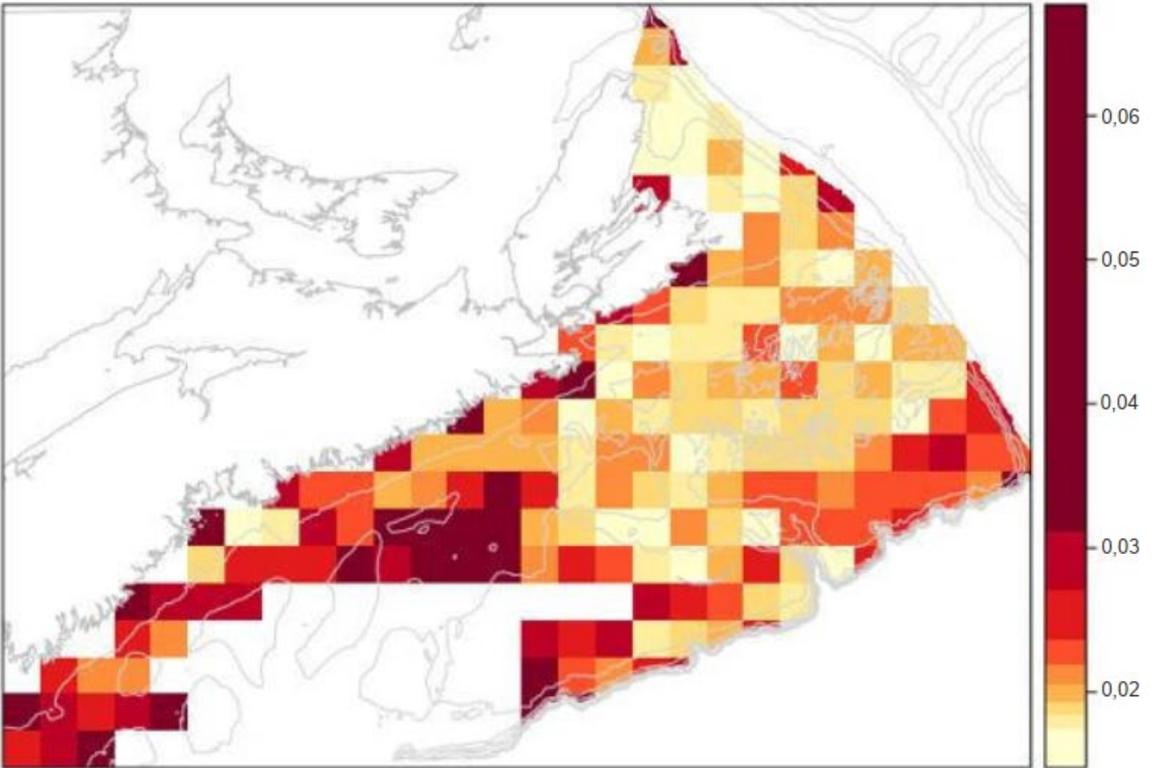
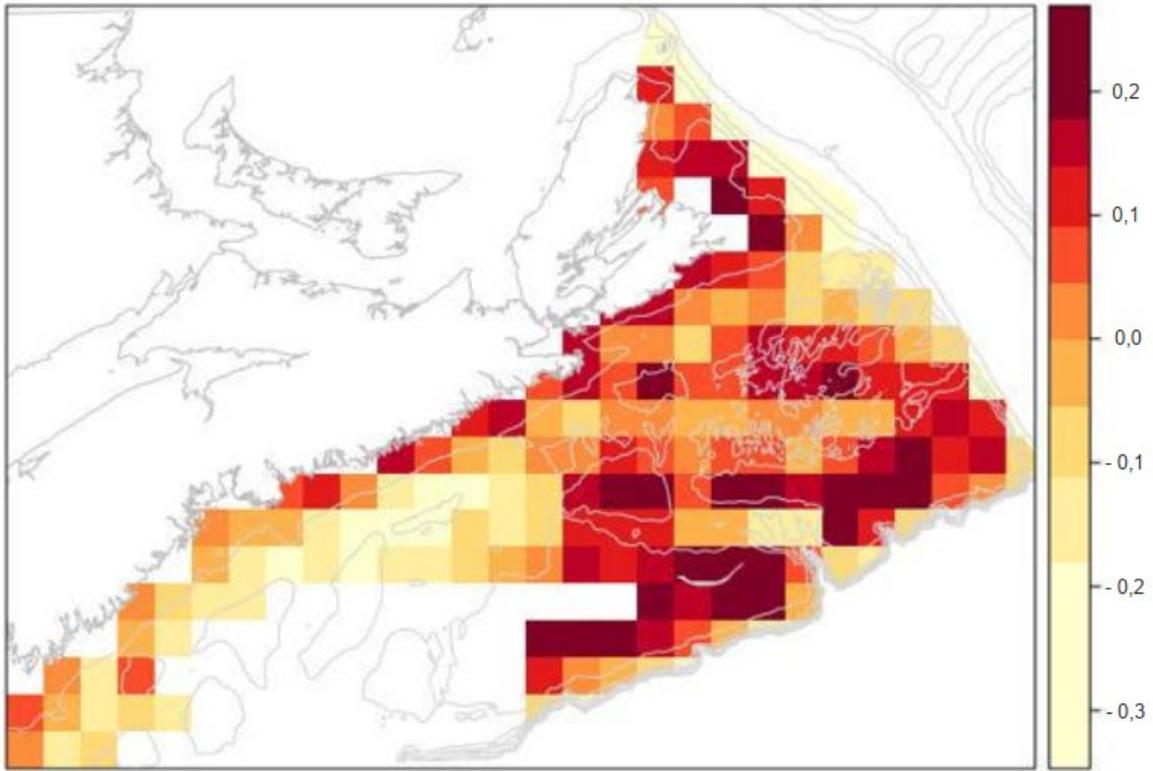


Figure 34. Axe 2 de composition de l'espèce pour septembre 2017; moyenne (en haut) et écart type (en bas).

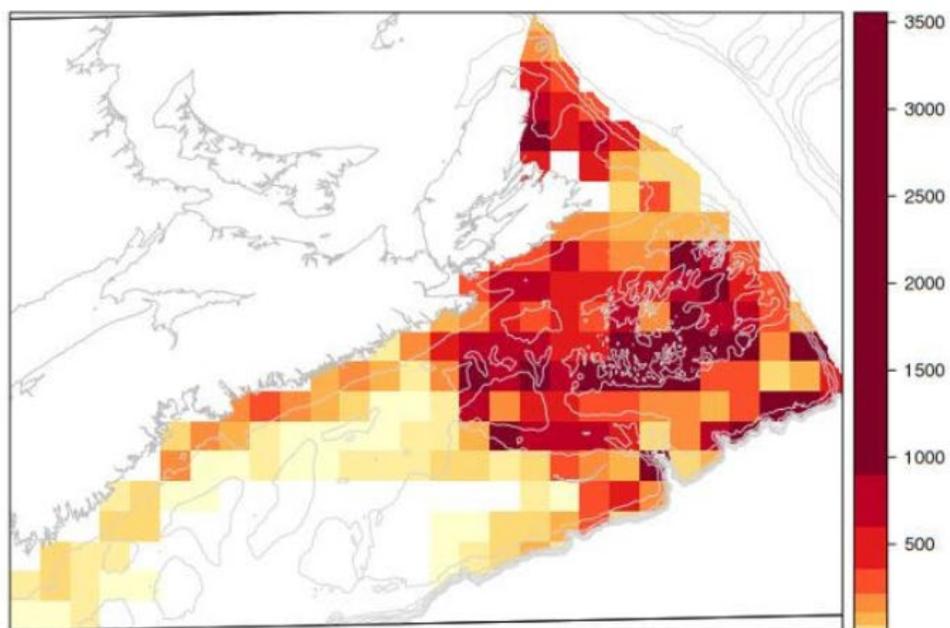
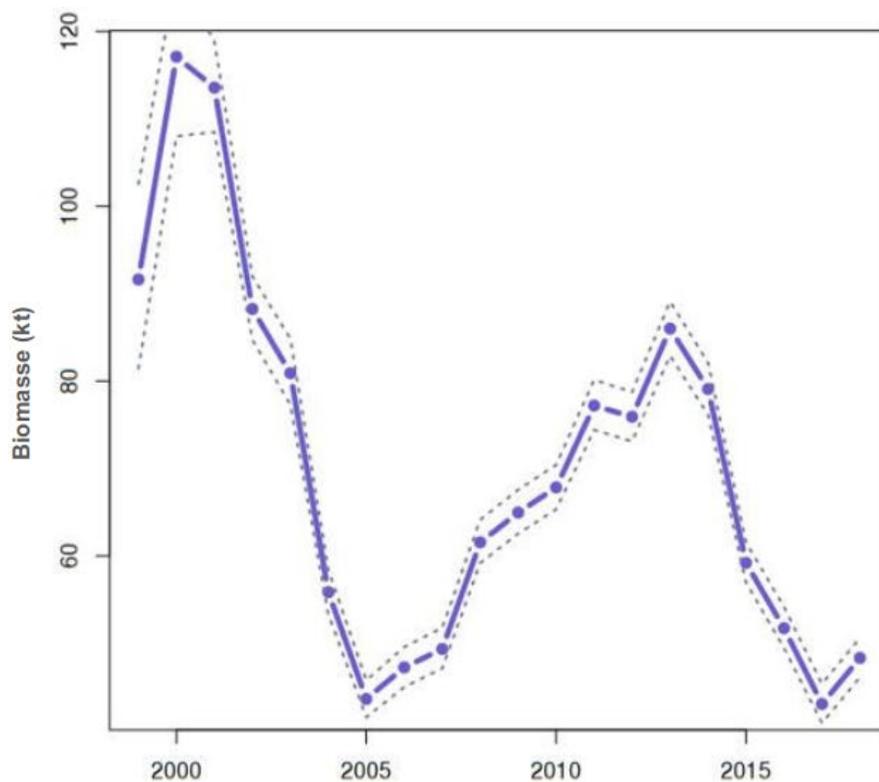


Figure 35. Modèle « dynamique à effets mixtes » : variations de la biomasse (kt) au fil des années traitées comme une variable factorielle et en supposant un effet aléatoire CAR additif unique et la taille des grains du substrat, la profondeur, la température au fond et la composition de l'espèce comme covariables avec une hypothèse de répartition de Poisson (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle. Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

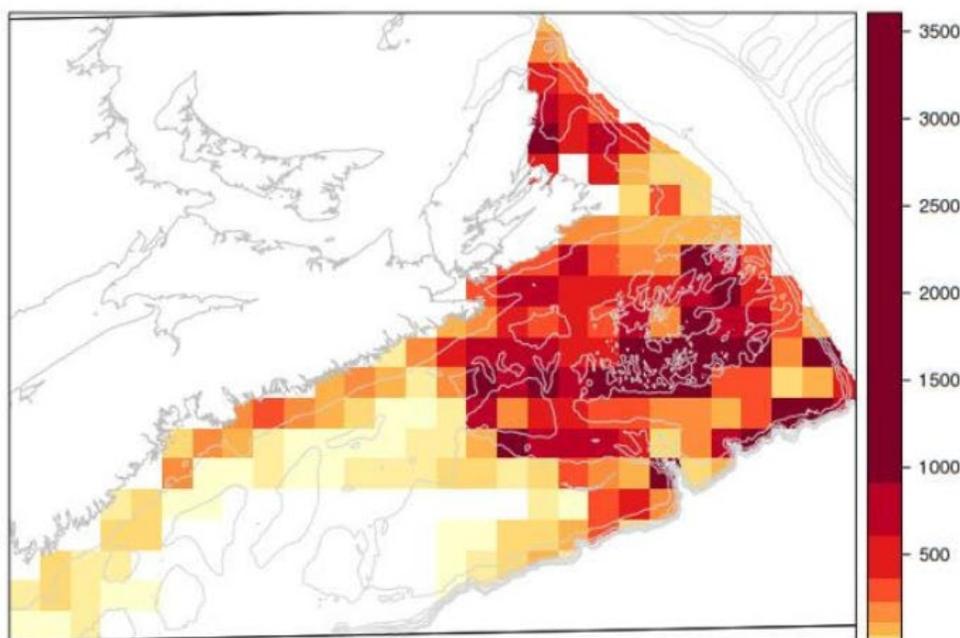
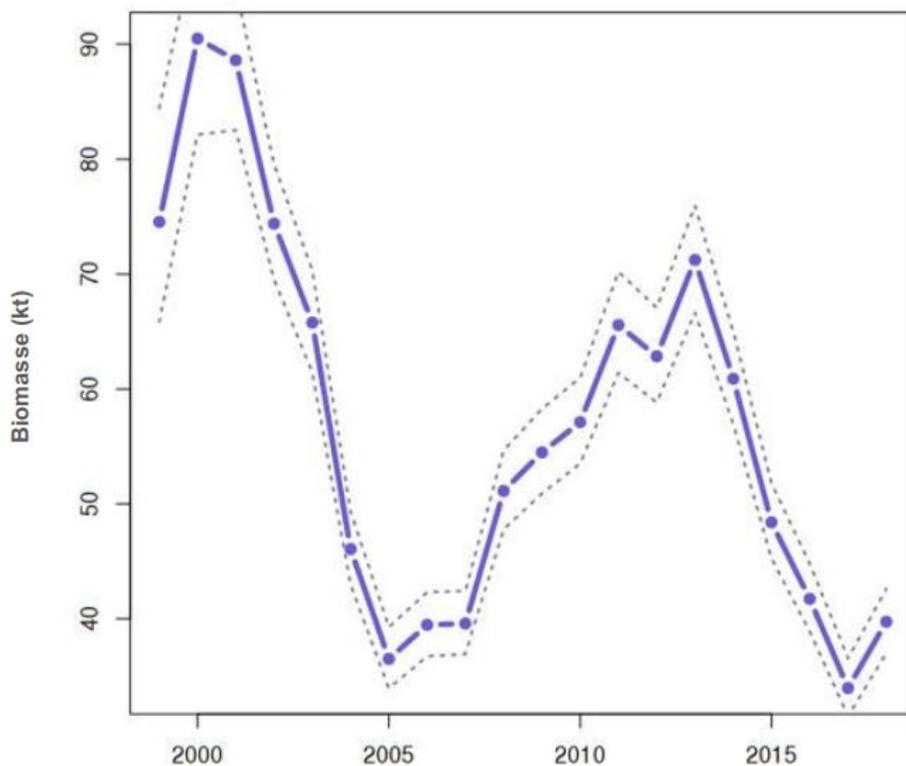


Figure 36. Modèle « séparable » : variations de la biomasse (kt) au fil des années traitées comme un processus AR1 et en supposant un effet aléatoire CAR additif unique et la taille des grains du substrat, la profondeur, la température au fond et la composition de l'espèce comme covariables avec une hypothèse de répartition de Poisson (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle. Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

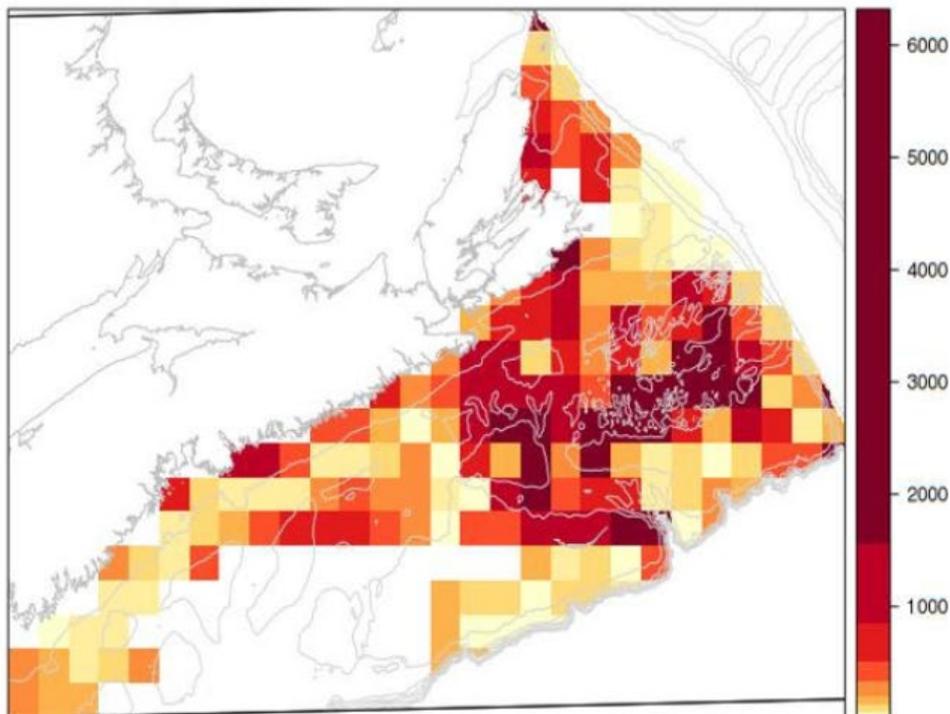
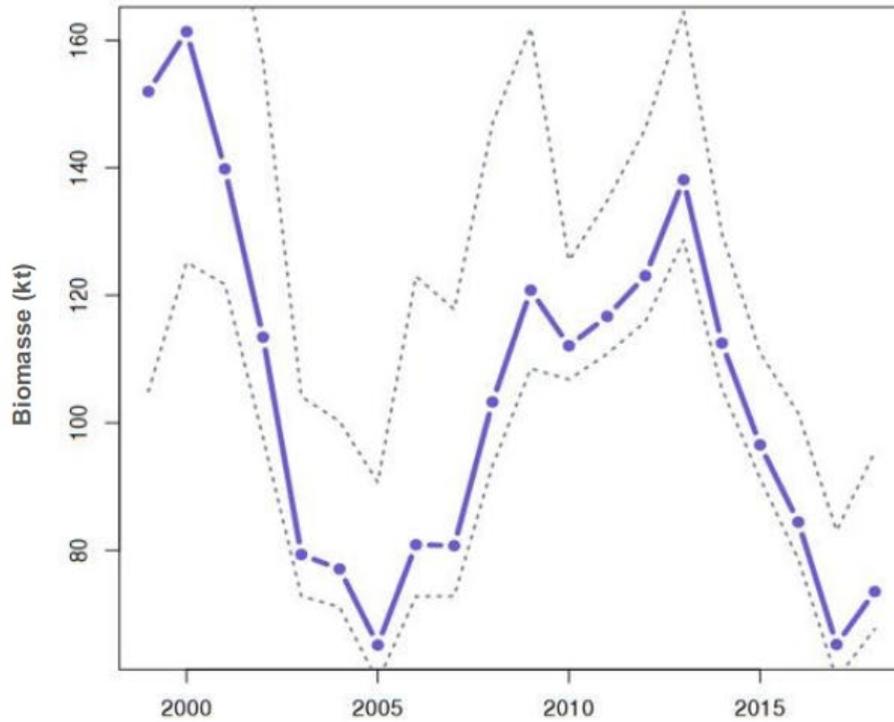


Figure 37. Modèle « **simple non séparable** » : variations de la biomasse (kt) au fil des années traitées comme un processus AR1 et en supposant un effet aléatoire CAR pour chaque année, aucune covariable avec une hypothèse de répartition de Poisson (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle. Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

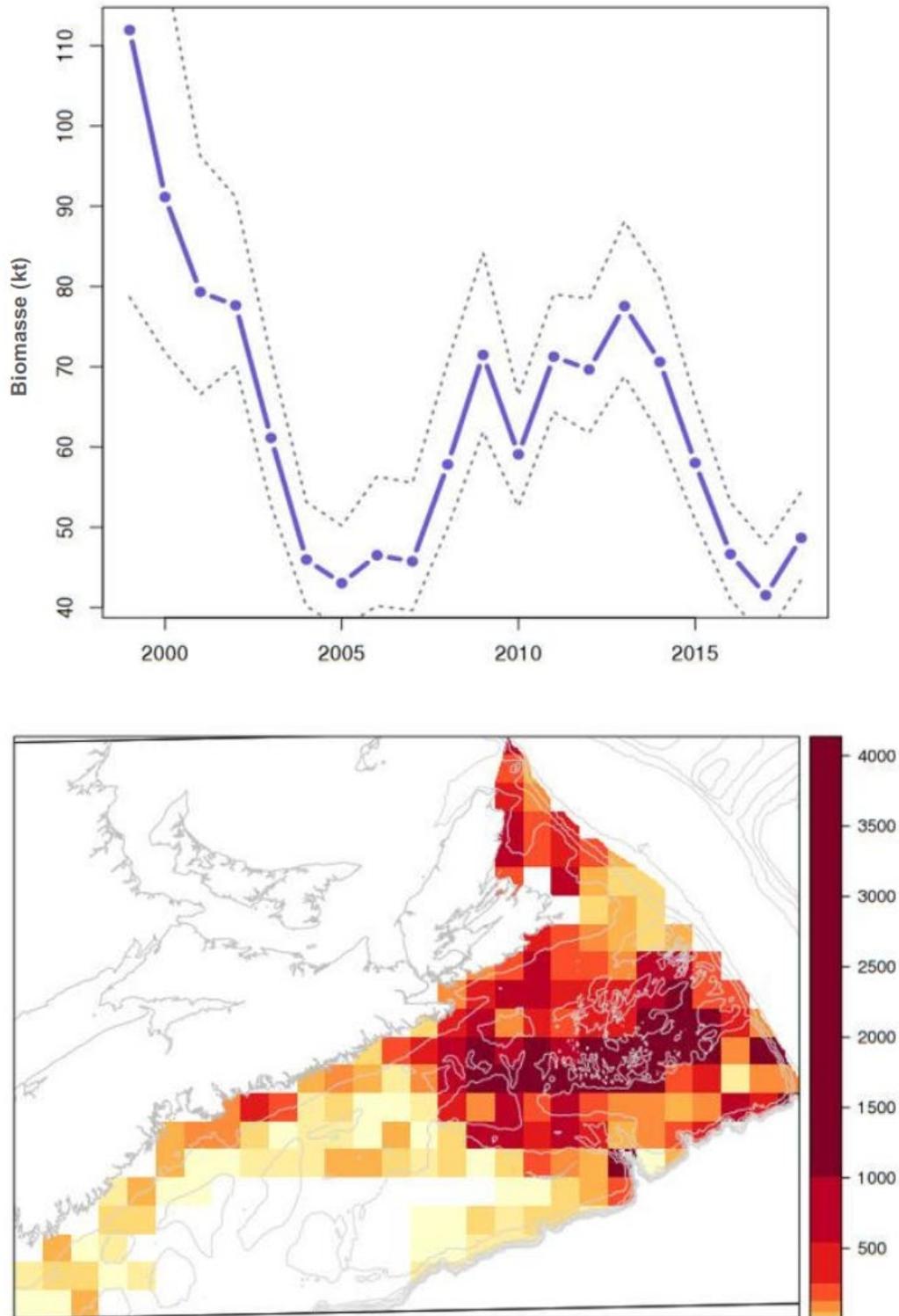


Figure 38. Modèle « **espace/temps non séparable** » : variations de la biomasse (kt) au fil des années traitées comme un processus AR1 et en supposant un effet aléatoire CAR pour chaque année, avec toutes les covariables et une hypothèse de répartition de Poisson (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle. Estimations de la biomasse connexes dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile des densités spatiales en kg/km^2).

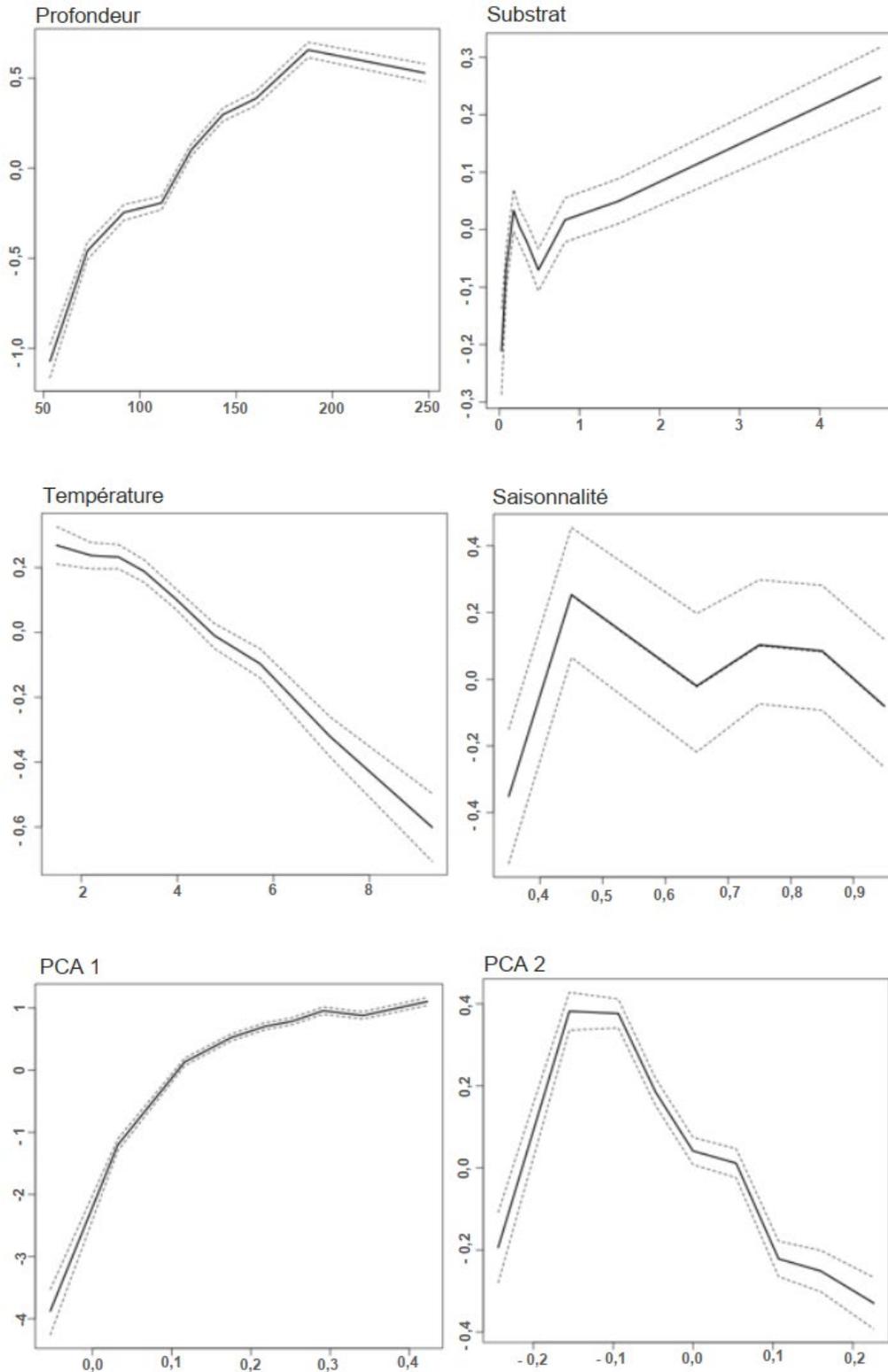


Figure 39. Relations des covariables avec la densité numérique du modèle « espace/temps non séparable ». Les covariables sont : profondeur (en haut à gauche), substrat (en haut à droite), température (au centre à gauche), saisonnalité (au centre à droite), PCA1 (en bas à gauche) et PCA2 (en bas à droite).

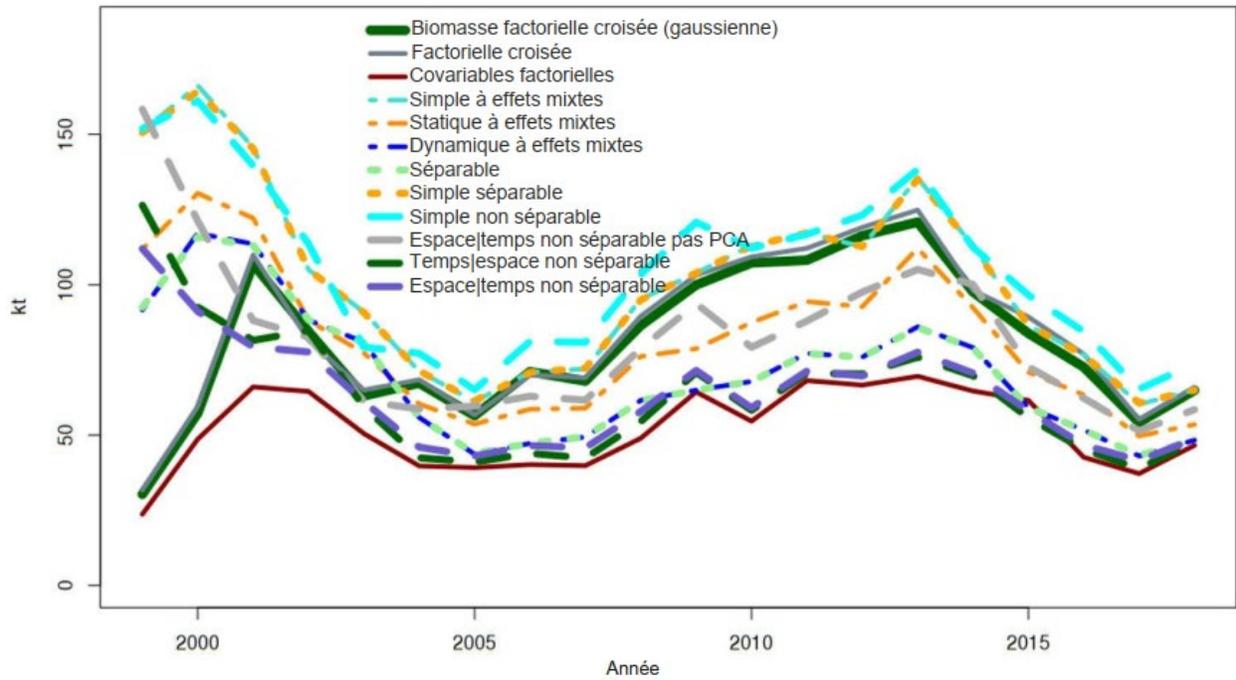


Figure 40. Comparaison des trajectoires de la biomasse $\int y_{st} ds$ pour tous les modèles pour le domaine complet de la région des Maritimes (kt).

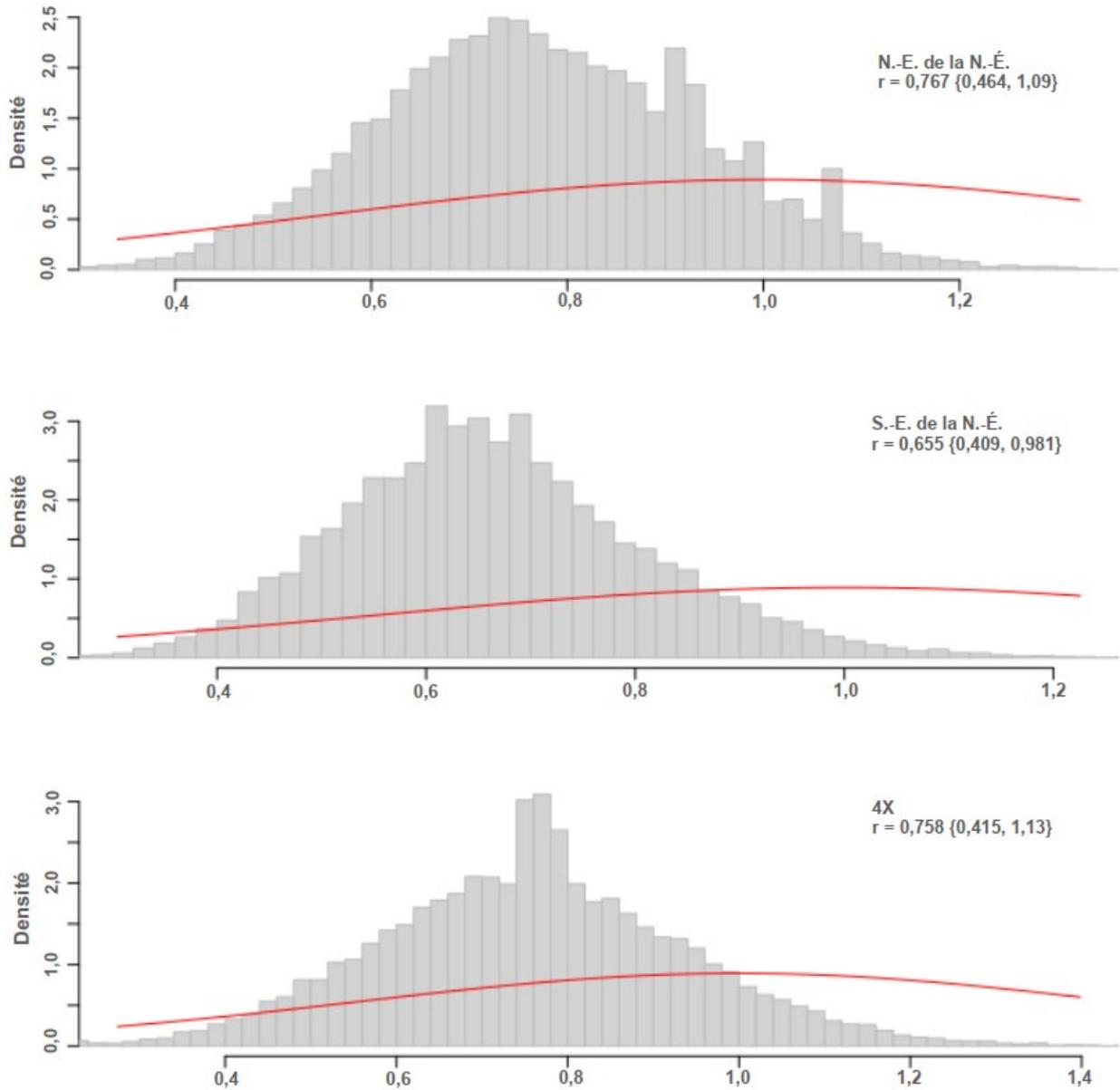


Figure 41. Distribution a posteriori du taux d'augmentation intrinsèque avec l'a priori en rouge.

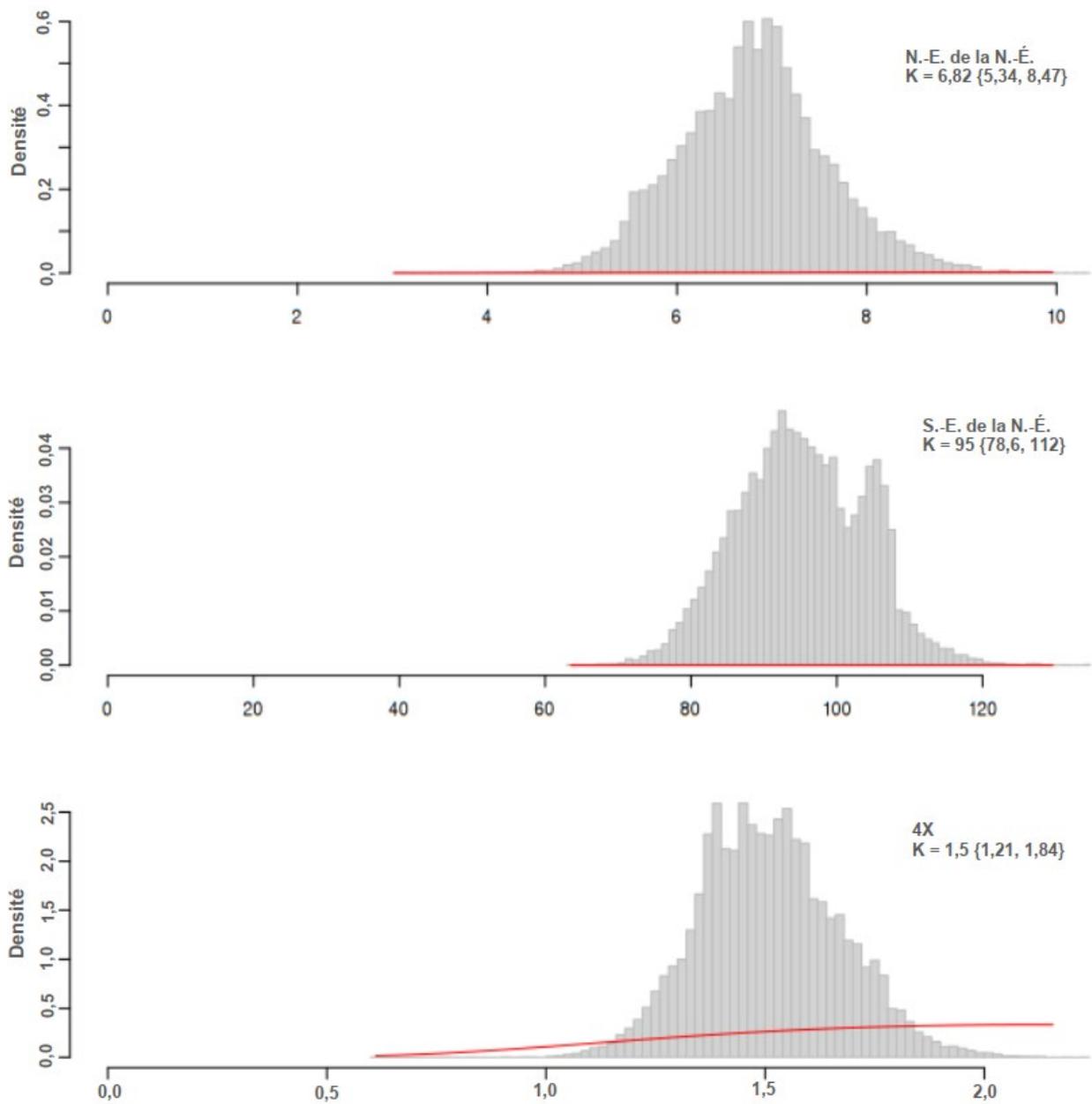


Figure 42. Distribution a posteriori de la capacité de charge avec l'a priori en rouge.

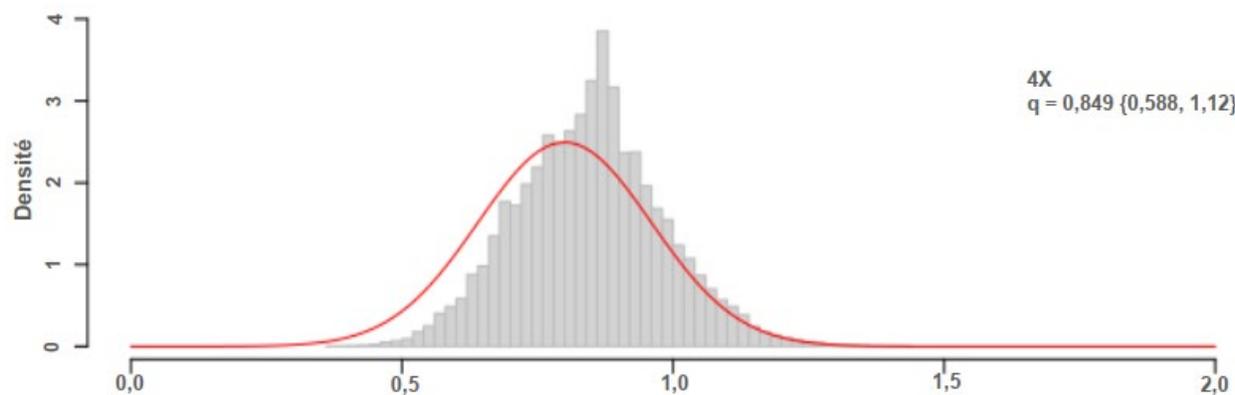
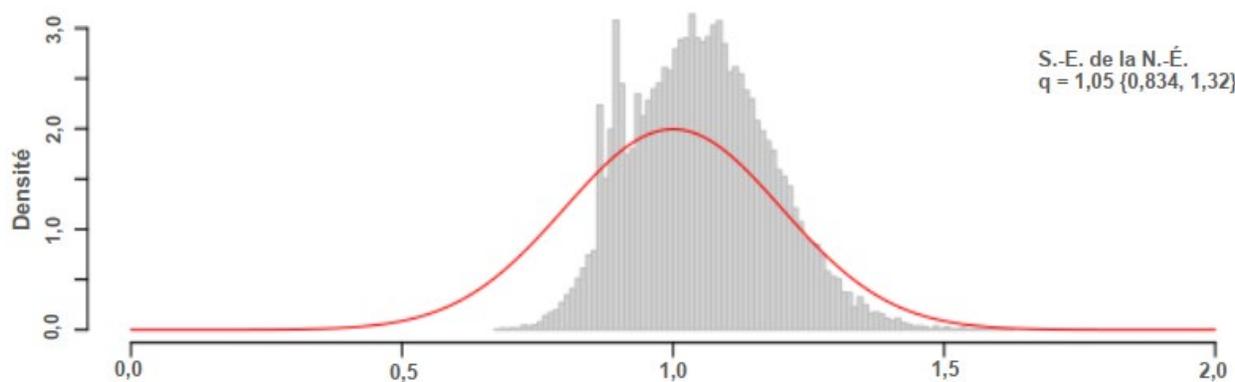
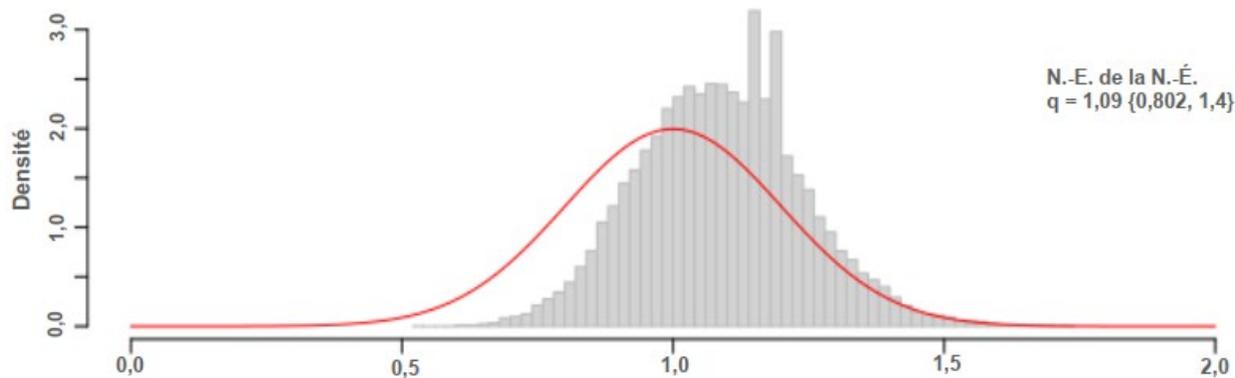


Figure 43. Distribution a posteriori du coefficient de capturabilité avec l'a priori en rouge.

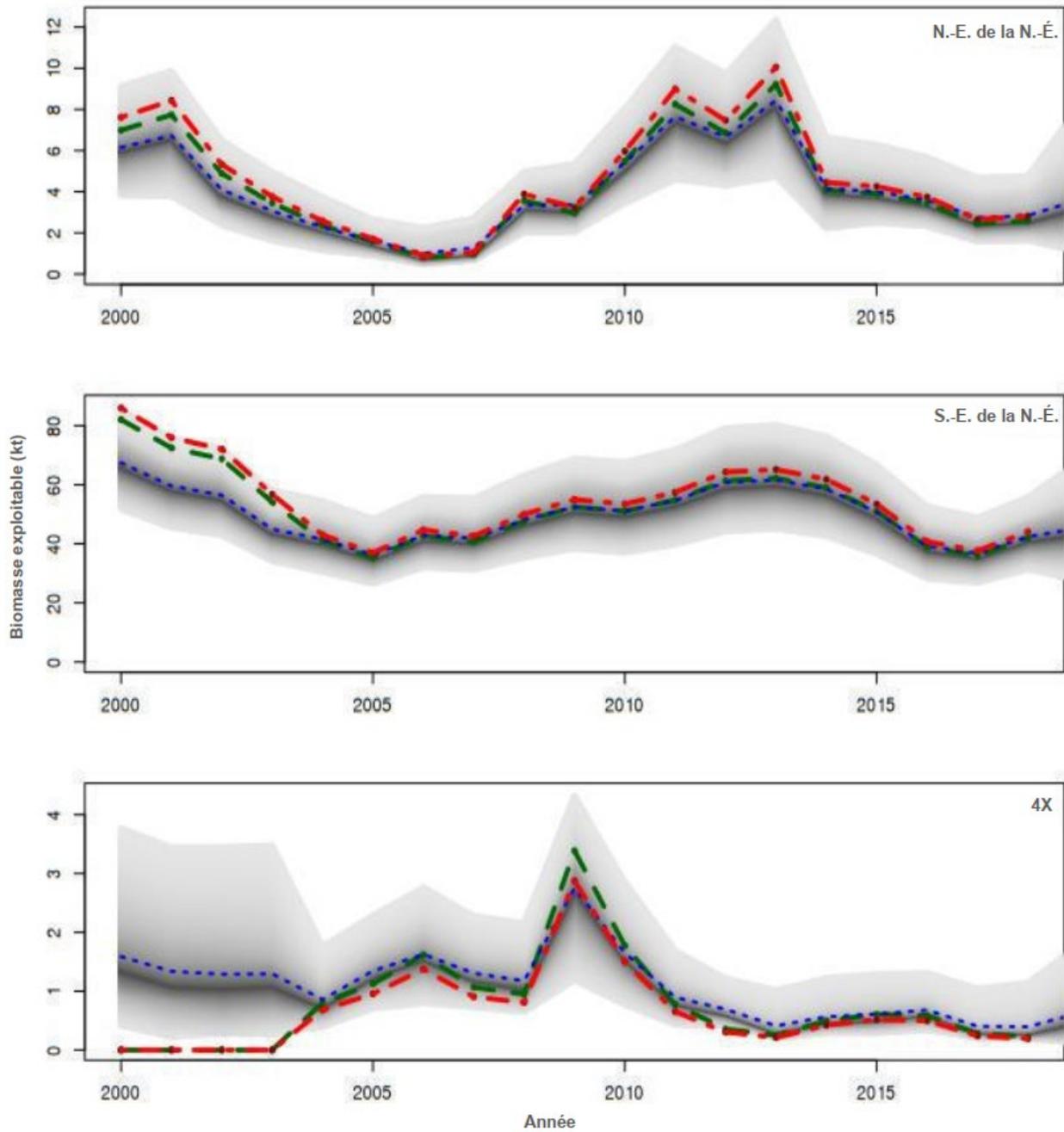


Figure 44. Comparaison des trajectoires de la biomasse. Le bleu correspond à la moyenne a posteriori de l'estimation de la biomasse Y_t , le gris foncé à la médiane et le gris pâle à l'intervalle de crédibilité bayésien (ICB) à 95 %. Le rouge correspond à l'indice de biomasse agrégé ($\int y_{st} ds$). Le vert correspond à la biomasse corrigée par la capturabilité ($q^{-1} \int y_{st} ds$).

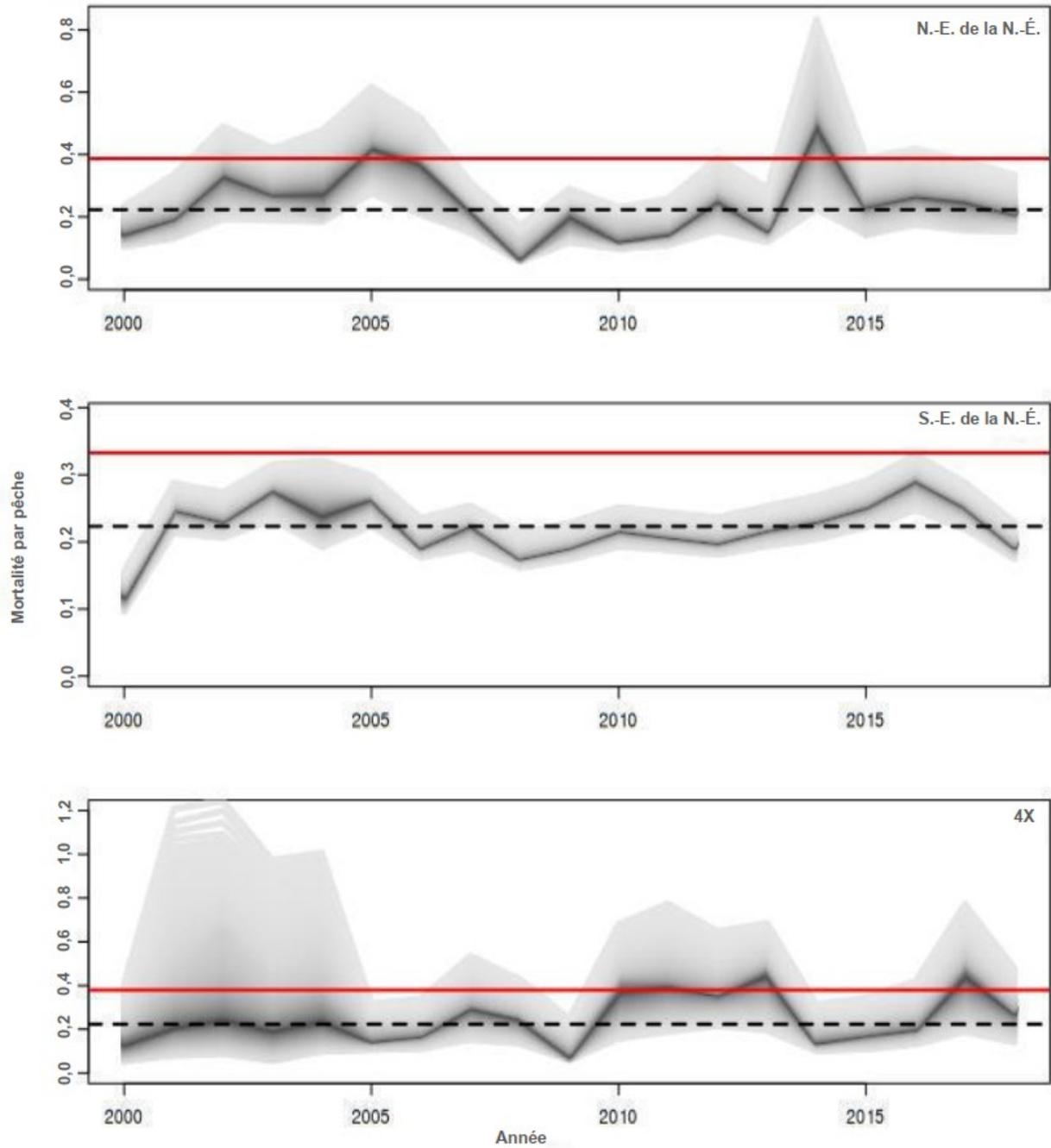


Figure 45. Comparaison des estimations de la mortalité par pêche. Le gris foncé correspond à la médiane et le gris pâle à l'intervalle de crédibilité bayésien à 95 %. La ligne rouge correspond à l'estimation de F_{RMD} . La ligne tiretée correspond au taux d'exploitation cible de 20 %.

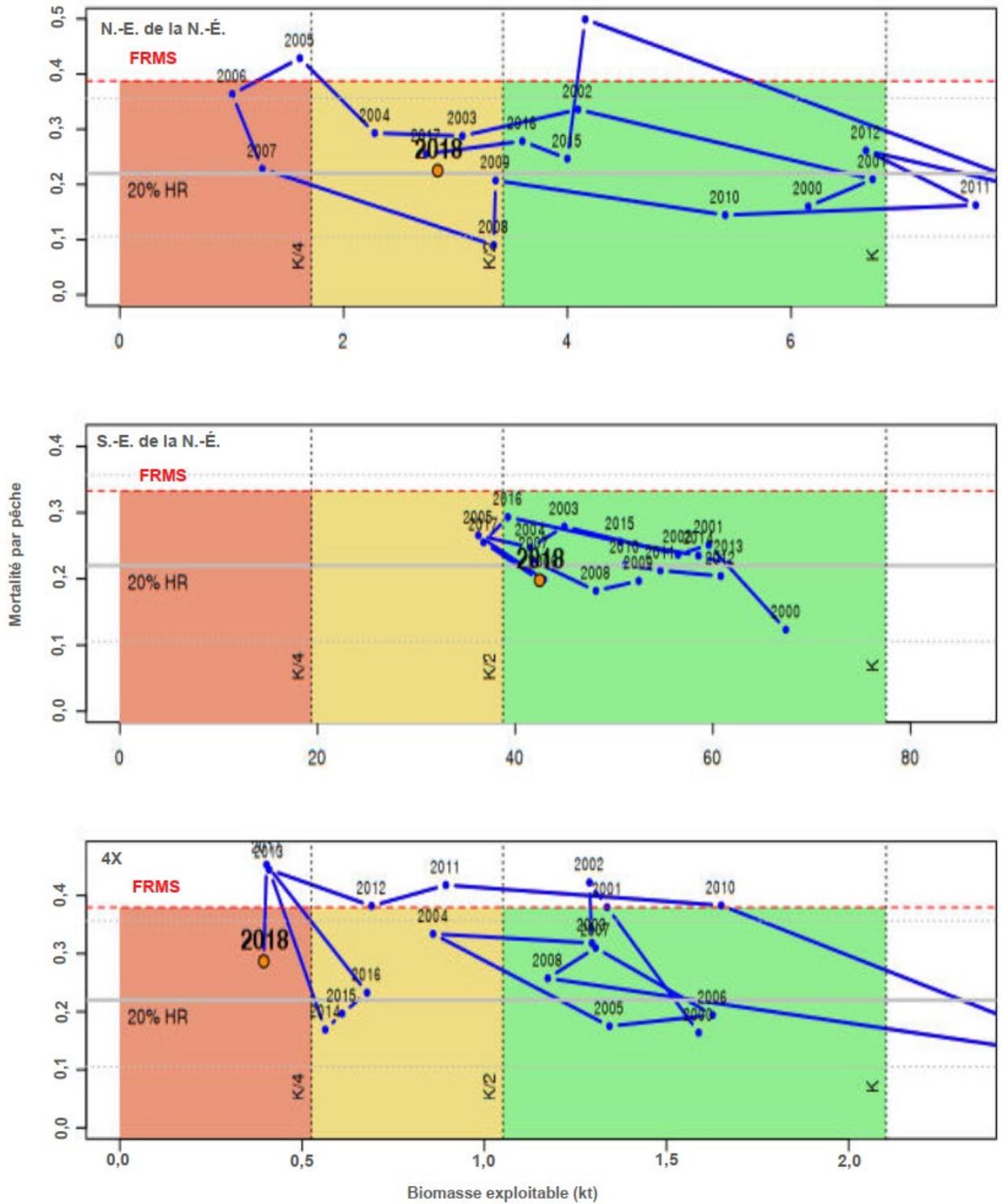


Figure 46. Règles de contrôle des prises.

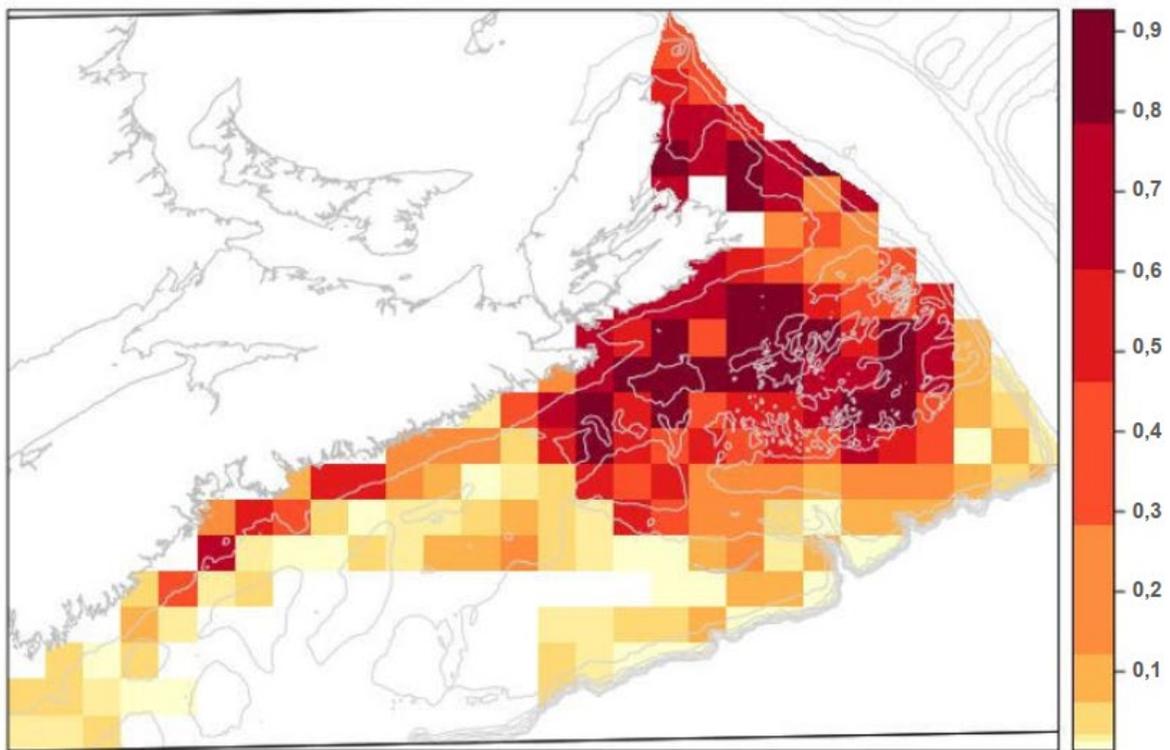
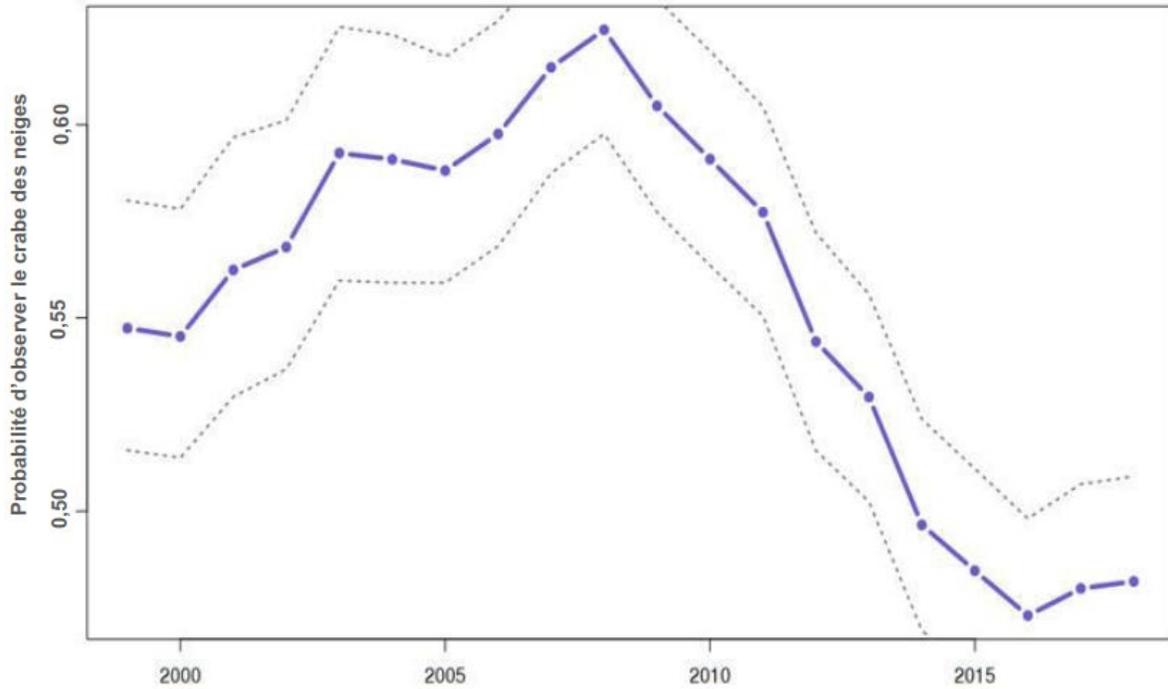


Figure 47. Modèle « habitat espace|temps non séparable » pour le petit crabe des neiges immature : probabilité d'observer le petit crabe des neiges pondérée par l'étendue de surface comme fonction du temps (en haut). L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle. Probabilités associées d'observer un petit crabe des neiges dans l'espace pour l'année 2017 (le rouge indique une valeur élevée et le jaune une valeur faible sur une échelle quantile de probabilité).

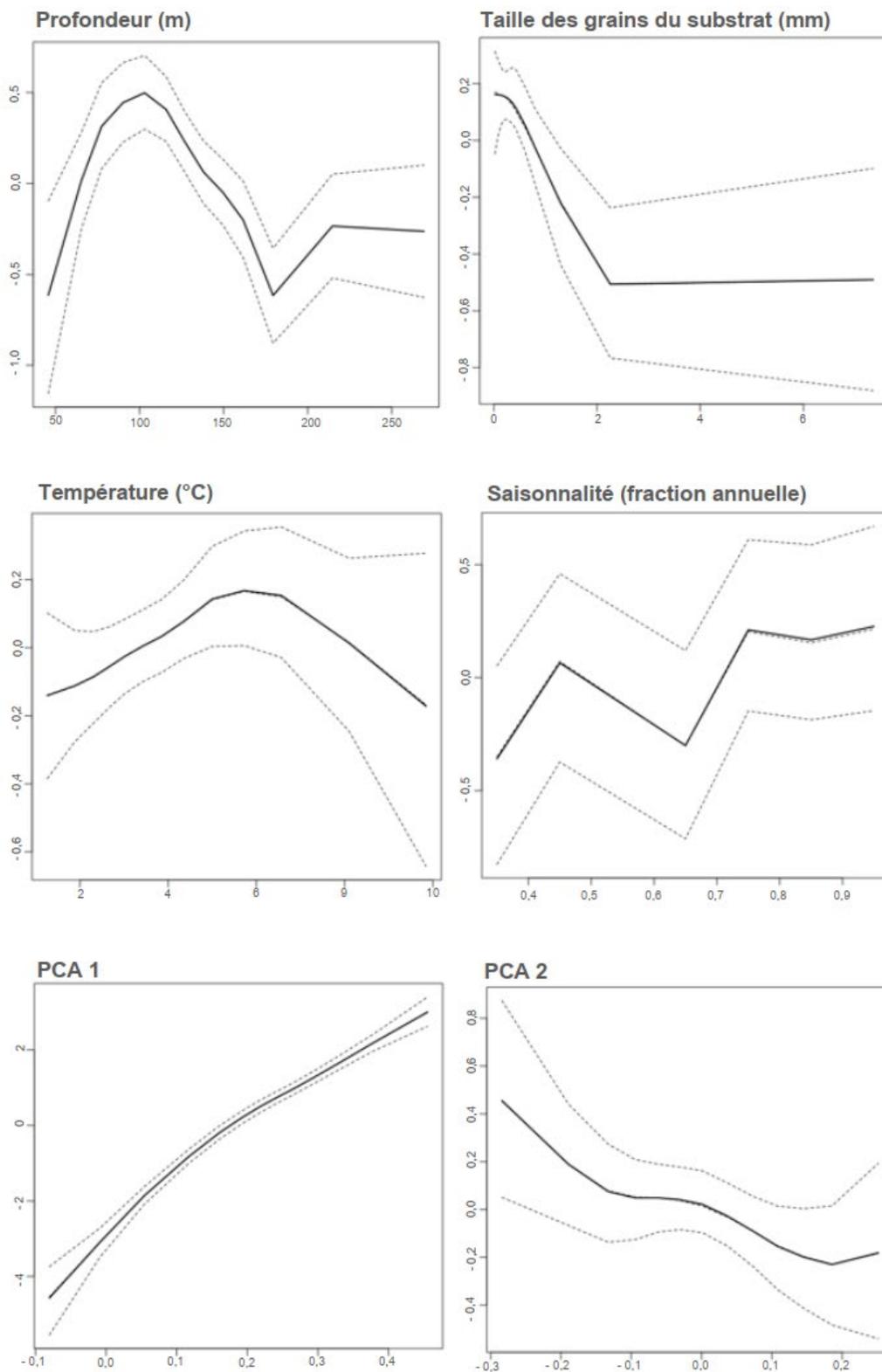


Figure 48. Relations des covariables avec la densité numérique pour le modèle « habitat espace|temps non séparable » du petit crabe des neiges immature (largeur de carapace <50 mm). Groupes : profondeur(m), taille des grains du substrat (mm), température (°C), saisonnalité (fraction annuelle), PCA1 et PCA2.

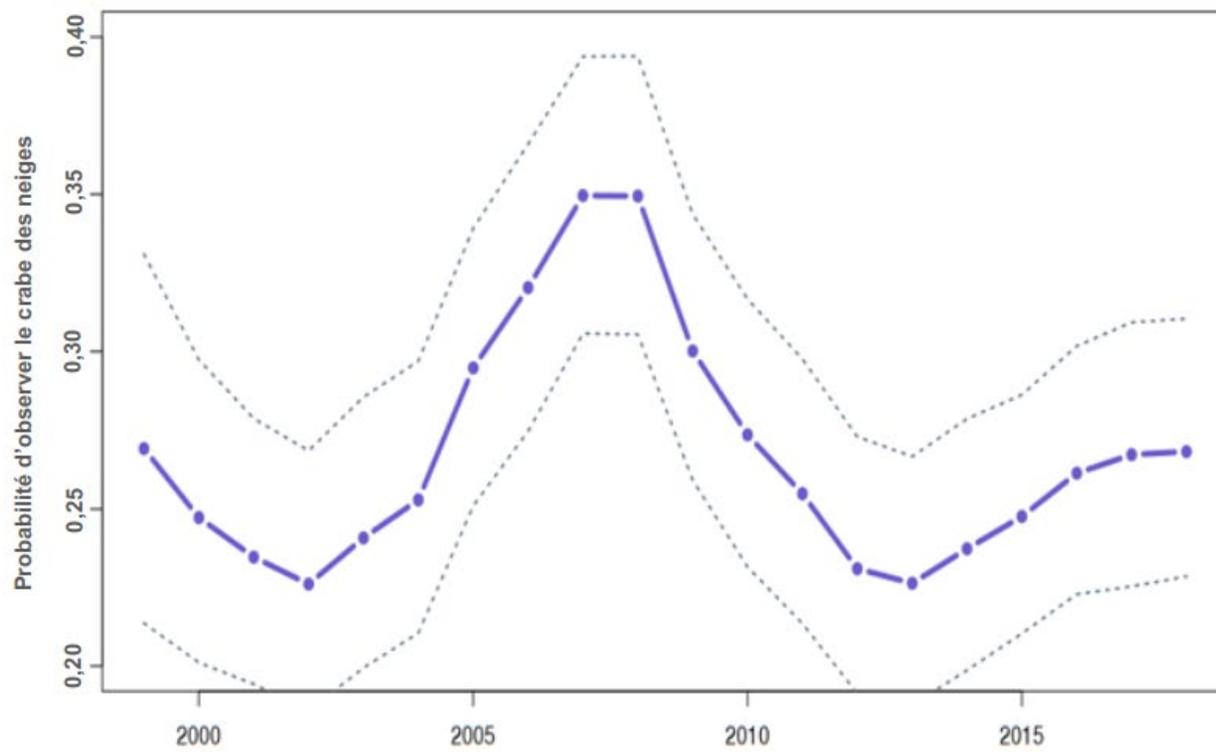
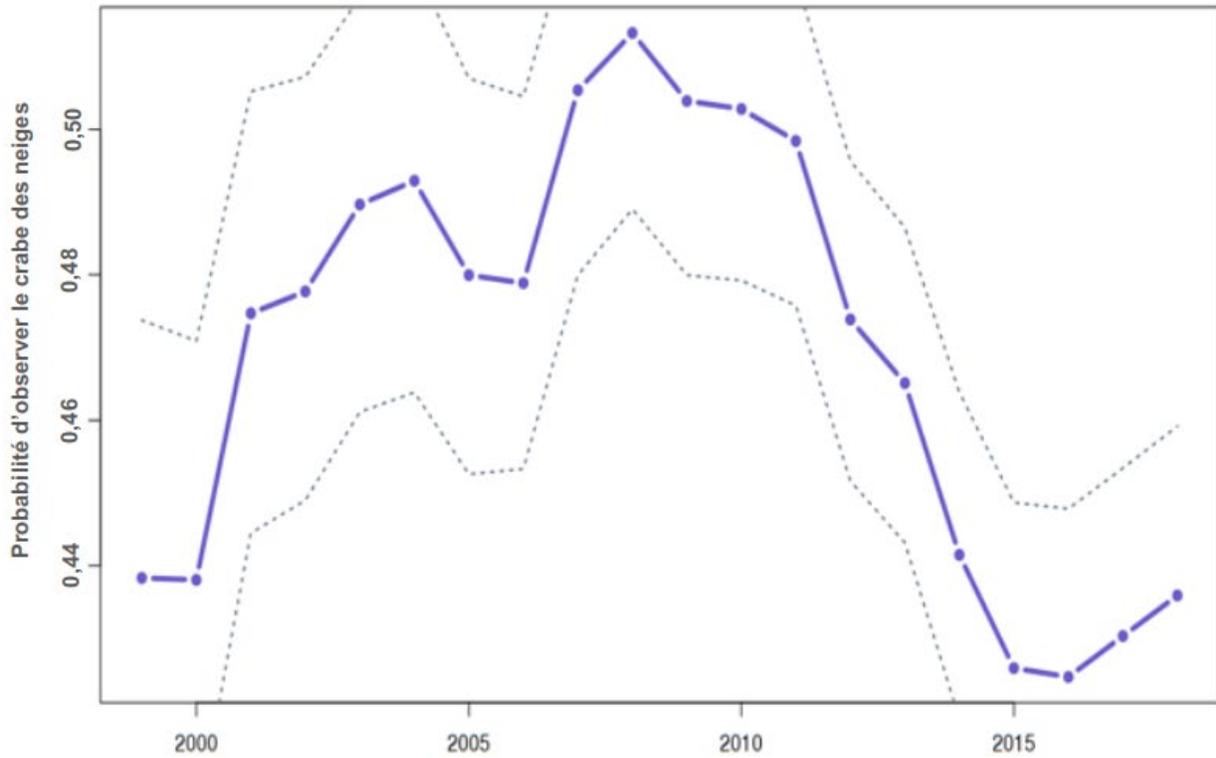


Figure 49. Modèle « habitat espace|temps non séparable » du crabe des neiges exploitable (en haut) et des femelles matures (en bas) : probabilité d'observer le crabe des neiges pondérée par l'étendue de surface comme fonction du temps. L'IC à 95 % repose sur des simulations a posteriori tirées des distributions conjointes des estimations du modèle.

RÉFÉRENCES CITÉES

- Banerjee, S., Carlin, B.P., and Gelfand, A.E.. 2004. Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability. Chapman and Hall/CRC.
- Bernardinelli, L., Clayton, D., and Montomoli, C. 1995. Bayesian estimates of disease maps: How important are priors? *Statistics in Medicine* 14: 2411–2431.
- Besag, Julian. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society Series B (Methodological)* 1974: 192–236.
- Besag, J., York, J., and Mollie, A. 1991. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math* 43: 1–20.
- Biron, M., Moriyasu, M., Wade, E., DeGrace, P., Campbell, R., and Hebert, M. 1997. [Assessment of the 1996 Snow Crab \(*Chionoecetes opilio*\) fishery off eastern Cape Breton, Nova Scotia \(Areas 20 to 24, and 4X\)](#). DFO Can. Sci. Advis. Sec. Res. Doc. 1997/102.
- Choi, J.S., Frank, K.T., Petrie, B., and Leggett, W.C. 2005a. Integrated assessment of a large marine ecosystem: A case study of the devolution of the eastern Scotian Shelf, Canada. *Oceanography and Marine Biology: An Annual Review* 43: 47–67.
- Choi, J.S., Zisserson, B., and Reeves, A. 2005b. [An assessment of the 2004 Snow Crab populations resident on the Scotian Shelf \(CFAs 20 to 24\)](#). DFO Can. Sci. Advis. Sec. Res. Doc. 2005/028.
- Choi, J.S. and Zisserson, B. 2011. [Assessment of Scotian Shelf Snow Crab in 2010](#). DFO Can. Sci. Advis. Sec. Res. Doc. 2011/110. viii + 95 p.
- Choi, J.S., Vanderlaan, A.S.M., Lazin, G., McMahan, M., Zisserson, B., Cameron, B. et Munden, J. 2008. [Cadre d'évaluation du banc de Sainte-Anne](#). Secr. can. de consult. scient. du MPO, Doc. de rech. 2018/066. vi + 68 p.
- Comeau, M., Conan, G.Y., Maynou, F., Robichaud, G., Therriault, J.C., and Starr, M. 1998. Growth, spatial distribution, and abundance of benthic stages of the Snow Crab (*Chionoecetes opilio*) in Bonne Bay, Newfoundland, Canada. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 262–279.
- Conan, G.Y., Comeau, M., and Robichaud, G. 1992. Life history and fishery management of Majid crabs: The case study of the Bonne bay (Newfoundland) *Chionoectes opilio* population. International Council for the Exploration of the Seas, Council Meeting Document 1992/K:21–21.
- Cooley, J.W., and Tukey, J.W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* 19 (90): 297–301.
- Cressie, N. 1993. *Statistics for spatial data*. Wiley-Interscience, New York.
- Danielson, G.C., and Lanczos, C. 1942. Some improvements in practical Fourier analysis and their application to x-ray scattering from liquids. *Journal of the Franklin Institute* 233: 365–380.
- DFO. 2006. [A harvest strategy compliant with the precautionary approach](#). DFO Can. Sci. Advis. Sec. Sci. Advis. Rep 2006/023.
- Elnor, R.W., and Beninger, P. 1995. Multiple reproductive strategies in Snow Crab, *Chionoecetes opilio*: Physiological pathways and behavioural plasticity. *Journal of Experimental Marine Biology and Ecology* 193: 93–112.

-
- Fotheringham, A.S., Brunson, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester.
- Fourier, J. 1822. Théorie analytique de la chaleur. Firmin Didot Père et Fils, Paris.
- Foyle, T., O'Dor, R., and Elner, R. 1989. Energetically defining the thermal limits of the Snow Crab. *Journal of Experimental Biology* 145: 371–393.
- Heideman, M.T., Johnson, D.H., and Burrus, C.S. 1985. Gauss and the history of the fast Fourier transform. *Archive for History of Exact Sciences* 34: 265–277.
- Hengl, T., Heuvelink, G.B.M., and Stein, A. 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* 120: 75–93.
- Hooper, R. 1986. A spring breeding migration of the Snow Crab, *Chionoectes opilio* (O. Fabr.), into shallow water in Newfoundland. *Crustaceana*. 50: 257–264.
- Hung, C.P., Chen, S.G., and Chen, K.L. (2004). "Design of an efficient variable-length FFT processor," 2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512), pp. II-833.
- Khintchine, Alexander. 1934. Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen* 109: 604–615.
- Kostylev, V., and Hannah, C. 2007. Process-driven characterization and mapping of seabed habitats. *Special Paper - Geological Association of Canada*. 47: 171–184.
- Kuhn, P., and Choi, J.S. 2011. Influence of temperature on embryo developmental cycles and mortality of female *Chionoectes opilio* (Snow Crab) on the Scotian Shelf, Canada. *Fisheries Research* 107: 245–252.
- Leroux Brian, G., Lei Xingye, and Breslow Norman. 2000. Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. *In*: Halloran, M. Elizabeth, Berry Donald editors. *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. New York, NY: Springer New York. pp. 179–191.
- Lindgren, F., and Rue, H. 2015. Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, 63 (19).
- Nychka, D., Furrer, R., Paige, J., and Sain, S. 2017. [Fields: Tools for spatial data. R package version 9.8-3](#). Accédé en janvier 2020.
- Riebler, A., Sørbye, S.H., Simpson D., and Rue, H. 2016. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research* 25: 1145–1165.
- Robertson, C., and George, S.C. 2012. Theory and practical recommendations for autocorrelation-based image correlation spectroscopy. *Journal of biomedical optics*, 17, 080801.
- Sainte-Marie, B. 1993. Reproductive cycle and fecundity of primiparous and multiparous female Snow Crab, *Chionoectes opilio*, in the northwest Gulf of Saint Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences* 50: 2147–2156.
- Sainte-Marie, B., and Hazel, F. 1992. Molting and mating of Snow Crabs, *Chionoectes opilio*, in shallow waters of the northwest Gulf of St. Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences* 49: 1282–1293.

-
- Sainte-Marie, B., Raymond, S., and Brethes, J.-C. 1995. Growth and maturation of the benthic stages of male Snow Crab, *Chionoecetes opilio* (Brachyura: Majidae). *Canadian Journal of Fisheries and Aquatic Sciences* 52: 903–924.
- Sigrist, F., Künsch, H.R., and Stahel, W.A. 2012. A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. *Ann. Appl. Statist.* 6: 1452–1477.
- Simpson, D., Rue, H., Riebler, A., Martins, T.G., and Sørbye, S.H. 2017. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statist. Sci.* 32: 1–28.
- Sølna, K., and Switzer, P. 1996. Time trend estimation for a geographic region. *Journal of the American Statistical Association* 91: 577–589.
- Tobler, W. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–240.
- Watson, J. 1972. Mating behaviour of the spider crab, *Chionoecetes opilio*. *Journal of the Fisheries Research Board of Canada* 29: 447–449.
- Webb, J.B., Eckert G.L., Shirley, T.C., and Tamone, S.L. 2007. Changes in embryonic development and hatching in *Chionoecetes opilio* (Snow Crab) with variation in incubation temperature. *Biological Bulletin*. 213: 67–75.
- Wiener, N. 1930. Generalized Harmonic Analysis. *Acta Mathematica* 55: 117–258.
- Wikle, C.K., and Cressie, N. 1999. A dimension-reduced approach to space-time Kalman filtering. *Biometrika* 86: 815–829.
- Xu, K., Wikle, C.K., and Fox, N.I. 2005. A kernel-based spatio-temporal dynamical model for nowcasting weather radar reflectivities. *Journal of the American Statistical Association* 100: 1133–1144.
- Yates, F. 1937. The design and analysis of factorial experiments. *Technical Communication / Commonwealth Bureau of Soils*, 35: 95 p.

ANNEXE 1 : BIAIS D'ÉCHANTILLONNAGE

Il est bien entendu, selon le théorème central limite, que plusieurs échantillons sont nécessaires pour estimer la moyenne globale et que celle-ci converge avec un nombre infini d'échantillons. N'ayant pas le luxe de pouvoir échantillonner à l'infini, nous éprouvons des problèmes d'estimation dus à la petite taille des échantillons. Il existe des règles empiriques heuristiques selon lesquelles $n = 30$ suffit pour commencer à estimer une valeur stable de la tendance centrale. Dans le cas du crabe des neiges, chaque unité spatiale est échantillonnée dans 0 à 8 stations par année (figure A1.1).

Les propriétés de la limite centrale ne s'appliquent pas à des nombres aussi faibles. À l'aide d'une simple simulation d'une unité spatiale de 25 km x 25 km, avec des cellules discrétisées à 25 m x 25 m, un seul chalut de crabe des neiges correspond en moyenne à 0,0039 km² ce qui représente environ 6,25 cellules de la grille (figure A1.2, panneau supérieur) sur un total de 10⁶ cellules dans chaque unité spatiale. La simulation suggère que la variabilité de la moyenne ne se stabilise pas avant que > 200 cellules soient échantillonnées, soit environ 32 stations (figure A1.2, panneau inférieur). Par conséquent, la probabilité est forte qu'une moyenne fondée sur des échantillons aussi limités soit biaisée par rapport à la moyenne latente (vraie), surtout si on a effectué également un échantillonnage préférentiel. Cela confirme le problème d'utiliser des moyennes simples pour représenter une unité spatiale donnée; des renseignements supplémentaires sont nécessaires pour contraindre les estimations à des valeurs plus réalistes, ce qui implique l'utilisation d'une approche d'estimation fondée sur un modèle.

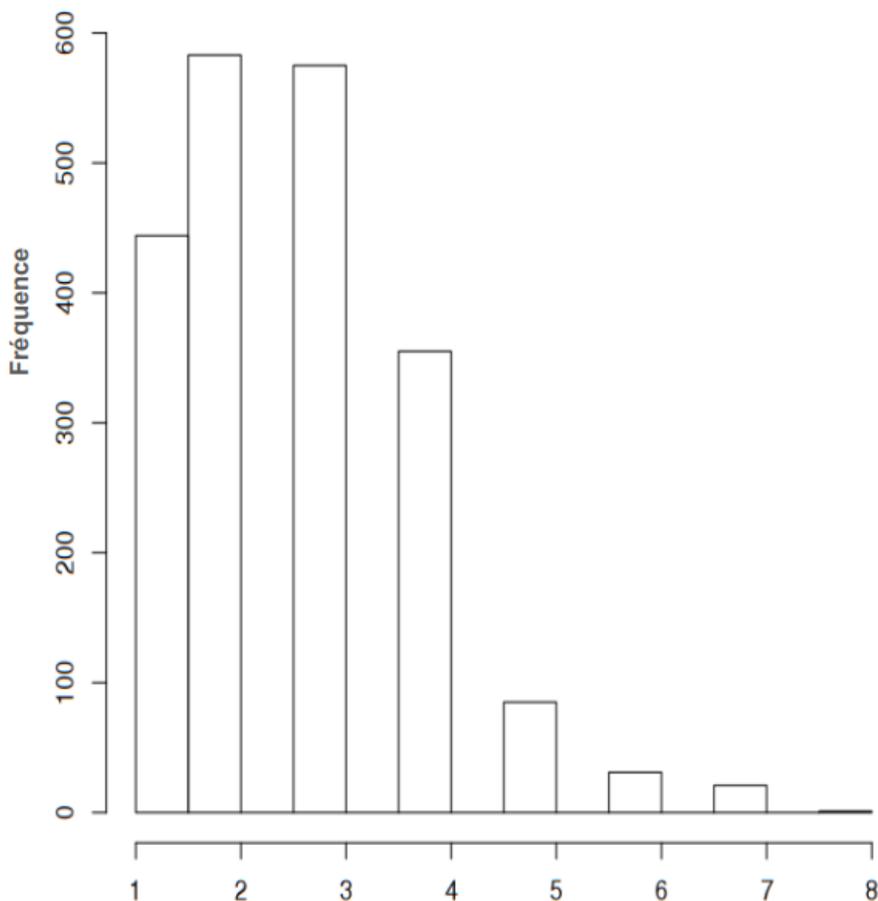


Figure A1.1. Nombre de stations dans chaque unité spatiale chaque année (moyenne = 2,6; écart type = 1,3).

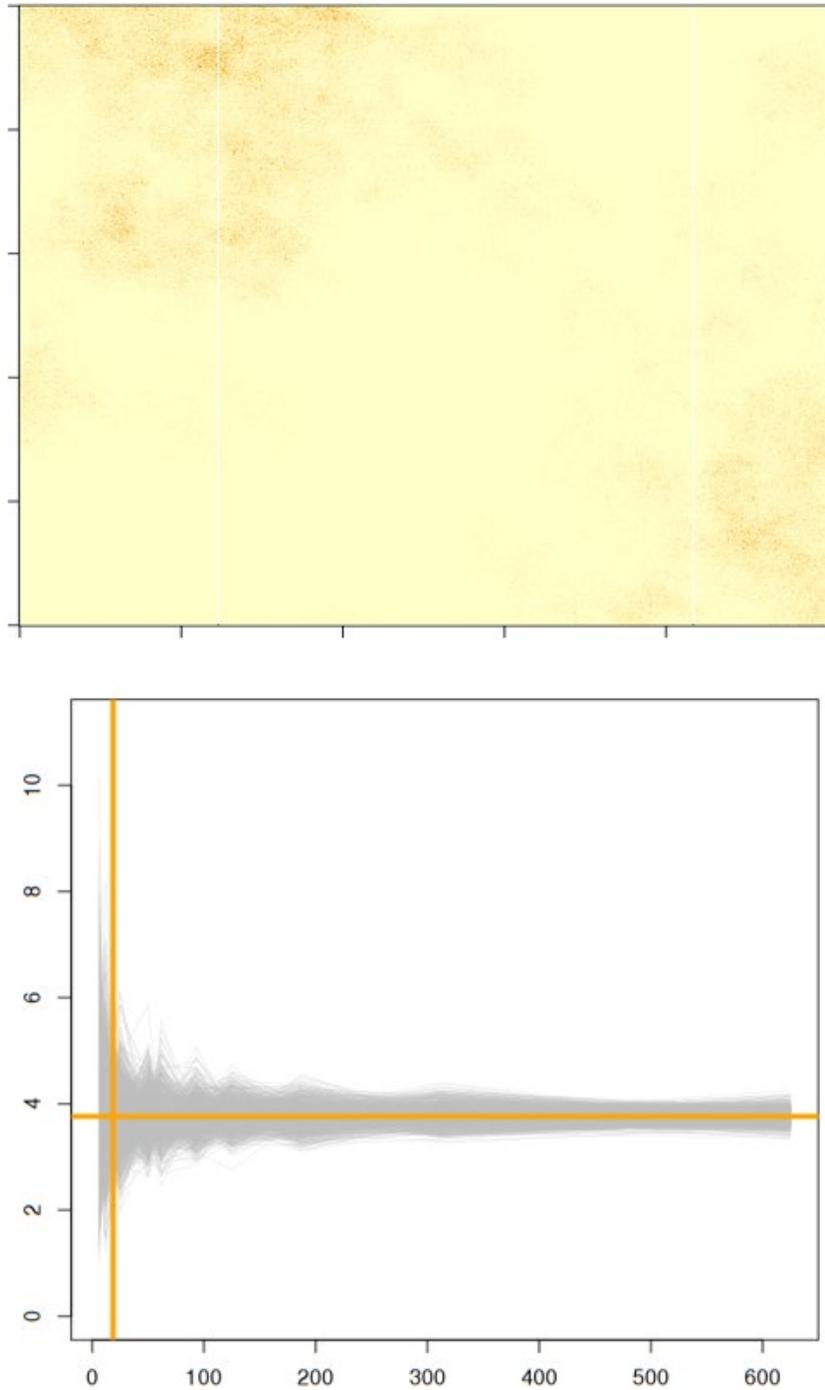


Figure A1.2. En haut : champ aléatoire gaussien, exponentialisé et tronqué pour imiter un processus de Poisson. Il a une échelle de variogramme exponentielle de 20 km (portée pratique) et une variance de 0,9 avec une variante indépendante des effets spatiaux de 0,1 et une moyenne globale de 1. Il a des dimensions de 25 km x 25 km et est discrétisé à 1000 unités dans chaque direction. Chaque unité de la grille mesure par conséquent 25 m x 25 m. La moyenne globale est de 5 individus par cellule de grille et l'écart type de 3,9. En bas : trajectoires des estimations moyennes comme fonction du nombre de cellules. Aux fins de référence, $n = 3$ stations (18,75 cellules) est donné comme la ligne verticale.

GRAPHIQUES DES RÉSIDUS

Les figures A1.3 à A1.11 correspondent aux résidus des modèles (observés-prévus) pour un sous-ensemble des modèles utilisés dans le présent document afin d'évaluer de manière approfondie la performance des modèles dans un contexte espace-temps et de compléter les mesures de performance des modèles DIC, WAIC et CPO, comme l'ont demandé les réviseurs du présent document. Toutes les méthodes montrent que l'hyperagrégation du crabe des neiges (rouge) et le faible nombre de crabes des neiges dans des zones où on s'attend à ce qu'il y en ait plus (bleu) présentent des difficultés. La faible abondance du crabe des neiges est probablement attribuable à l'épuisement localisé dans lesdites zones dû à une activité de pêche permanente. L'hyperagrégation est probablement due à de mauvaises conditions environnementales environnantes qui obligent le crabe à pénétrer dans des zones localisées où les conditions sont meilleures (« trous de crabe »). Dans les deux cas, les zones tendent à se situer dans le banc de Misaine, une zone de grande complexité topographique et de connectivité non simple.

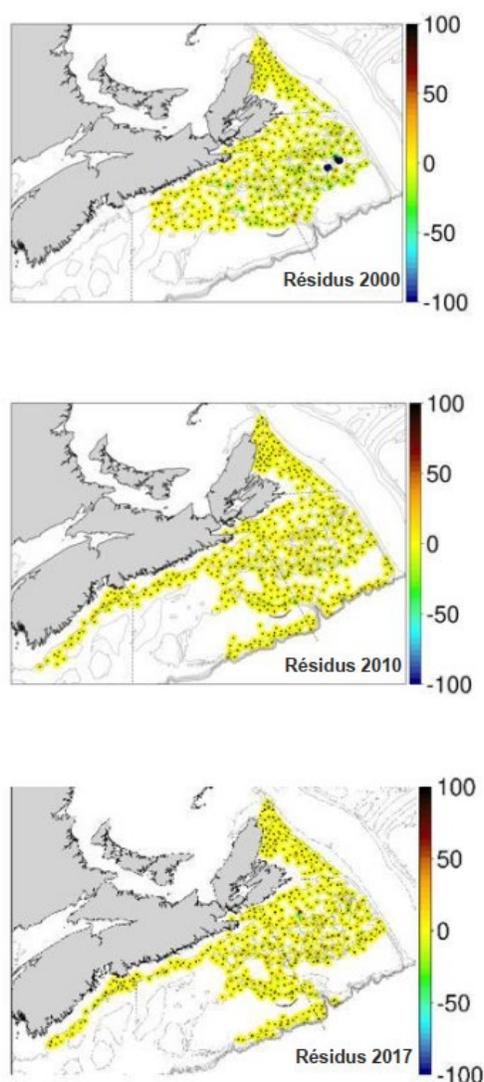


Figure A1.3. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges, sur l'échelle de densité de biomasse par ensemble (kg/calée)—**Biomasse factorielle croisée (gaussienne)**.

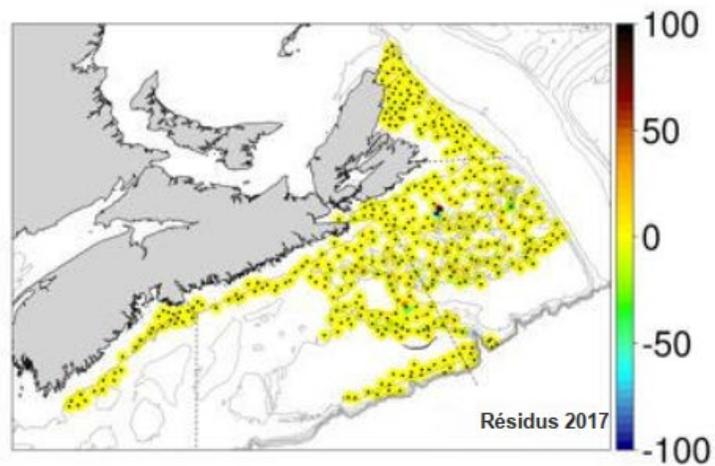
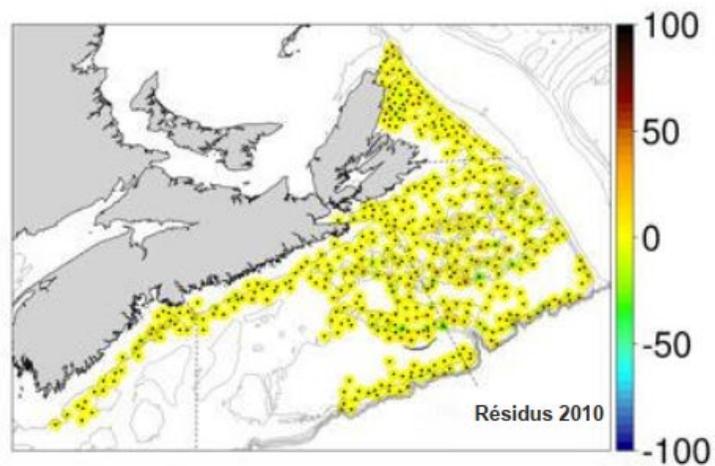
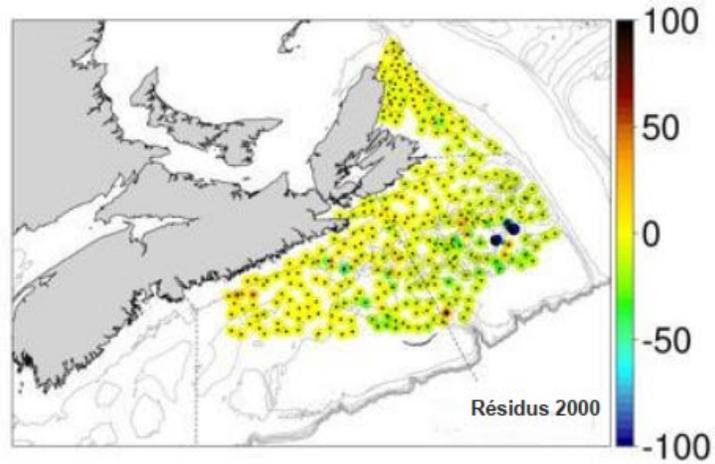


Figure A1.4. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée)— **Factoriel croisé**.

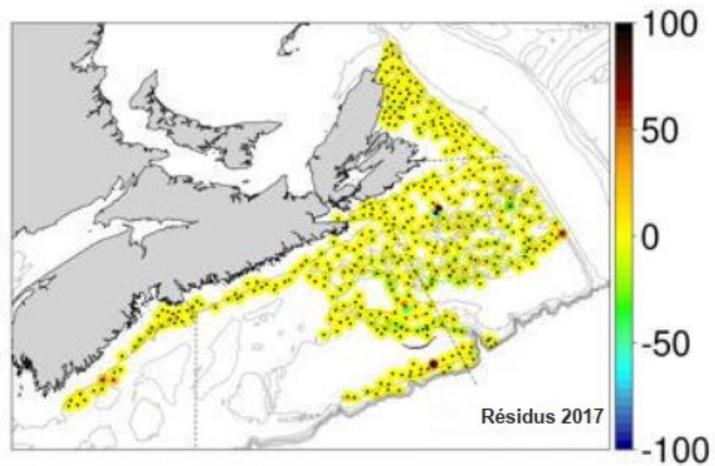
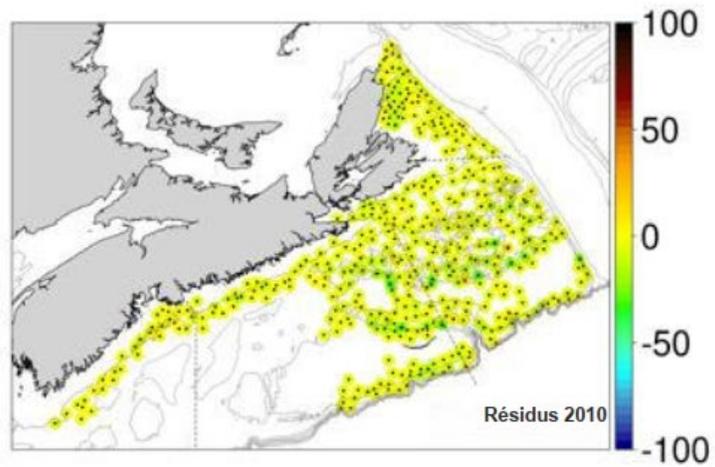
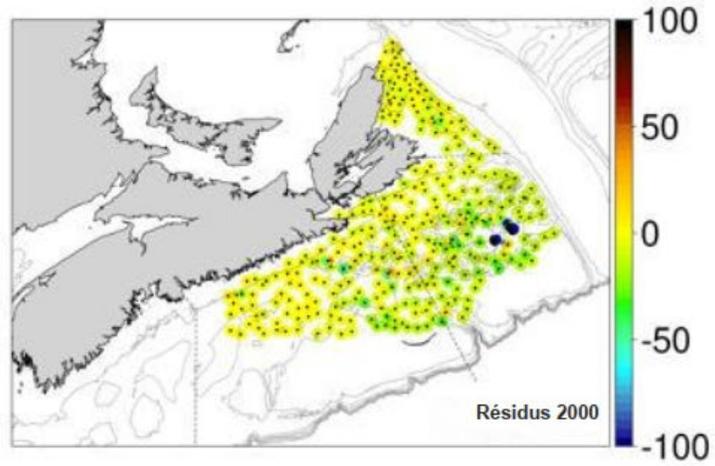


Figure A1.5. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée)—**Covariables factorielles.**

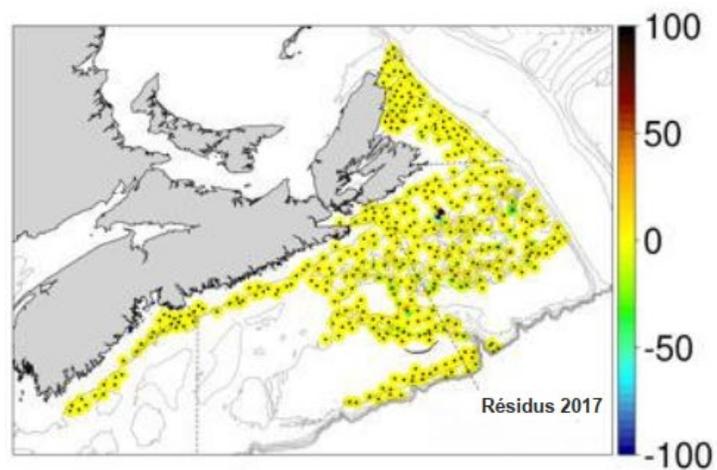
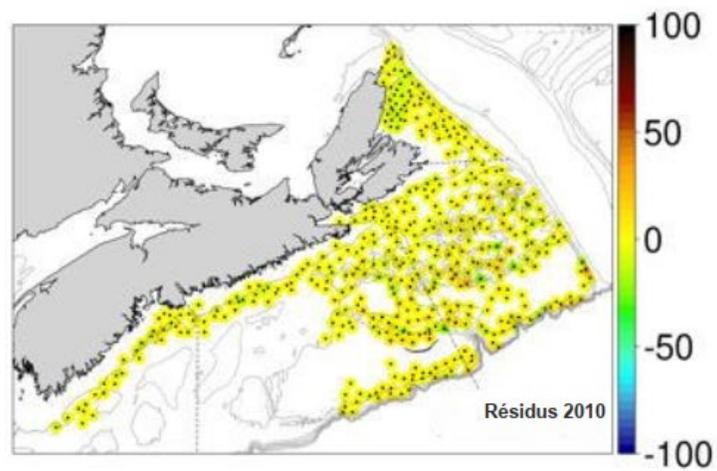
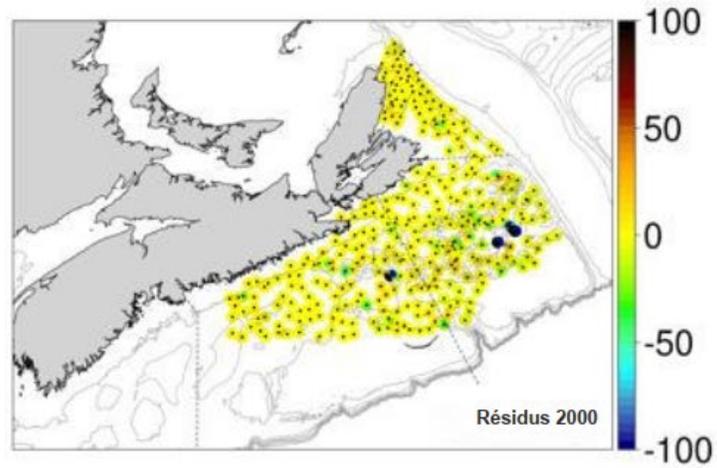


Figure A1.6. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée)—**Simple à effets mixtes**.

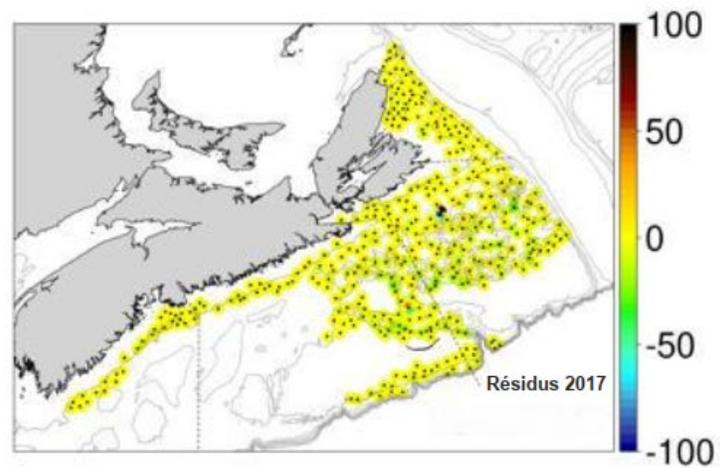
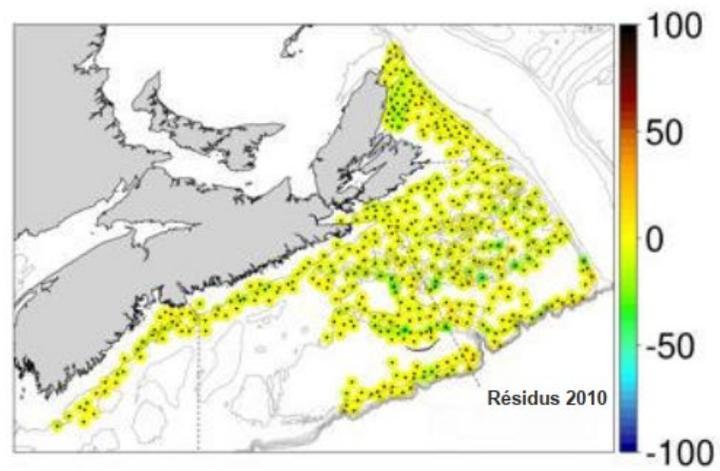
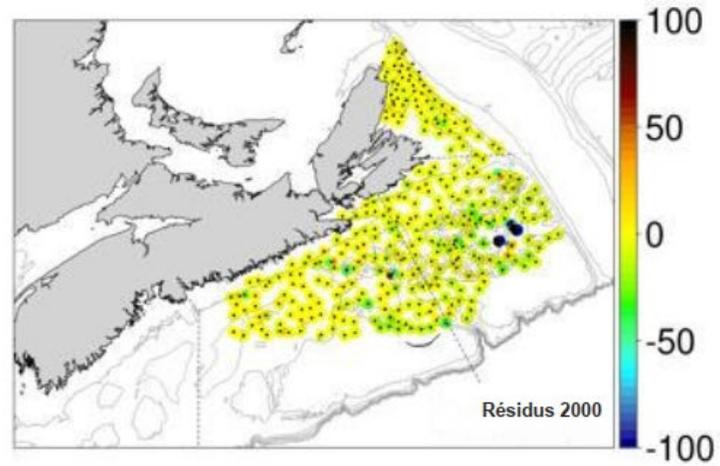


Figure A1.7. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée) – **Dynamique à effets mixtes.**

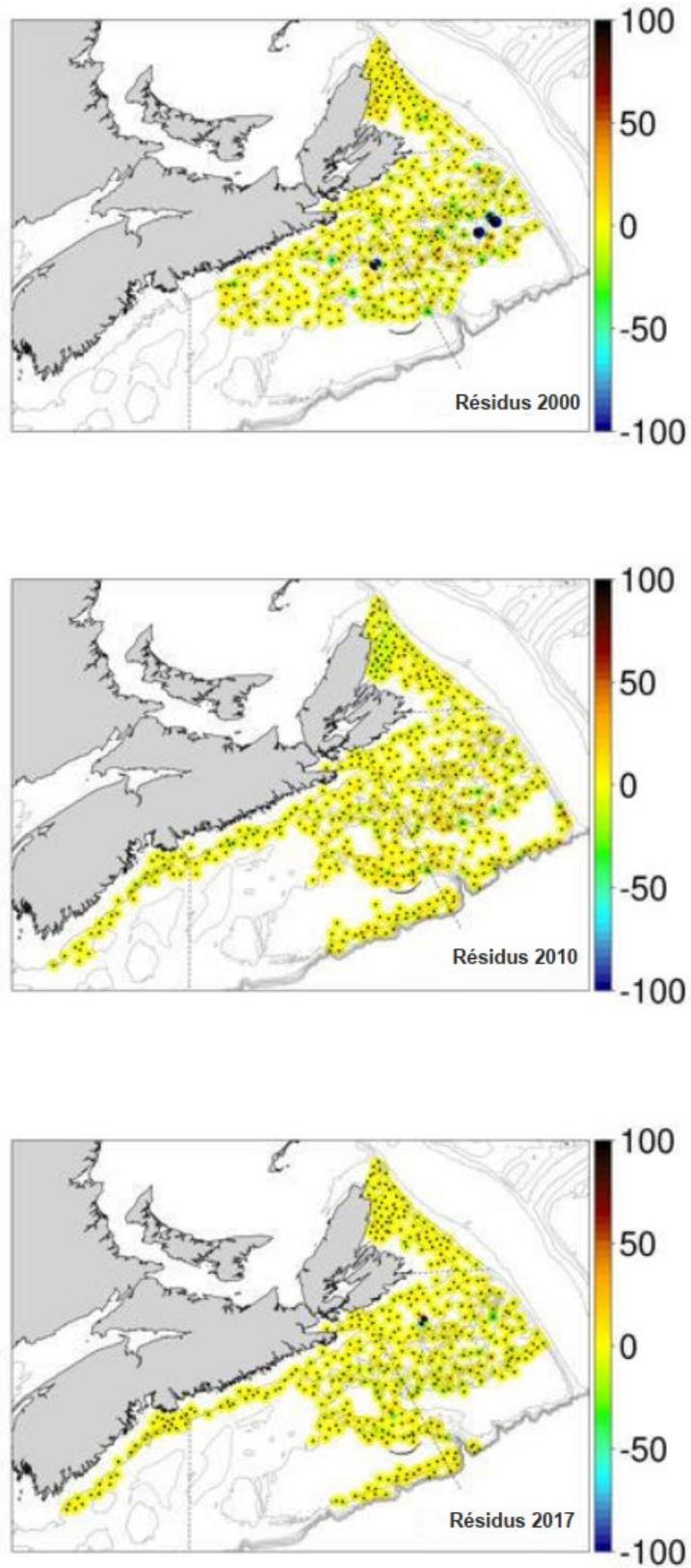


Figure A1.8. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée)—**Simple séparable**.

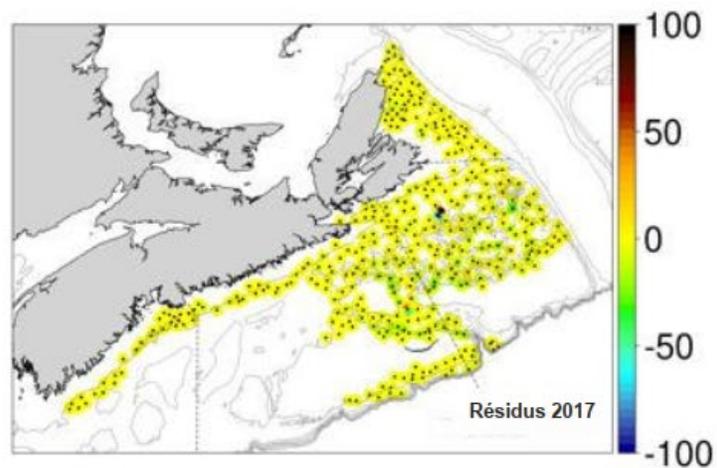
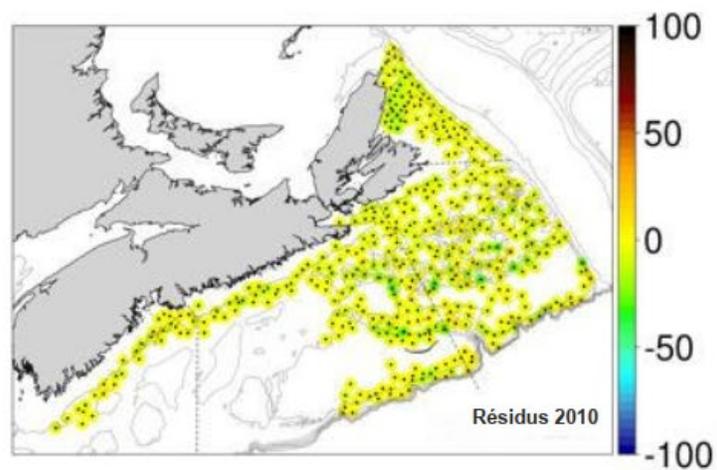
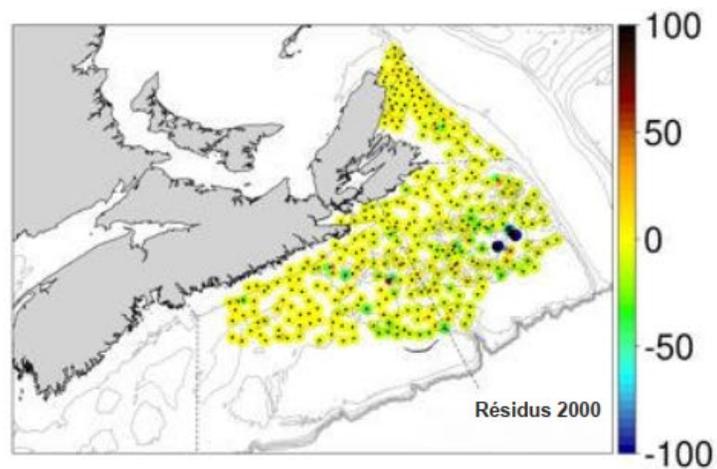


Figure A1.9. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges (nombre d'individus par calée)—**Séparable**.

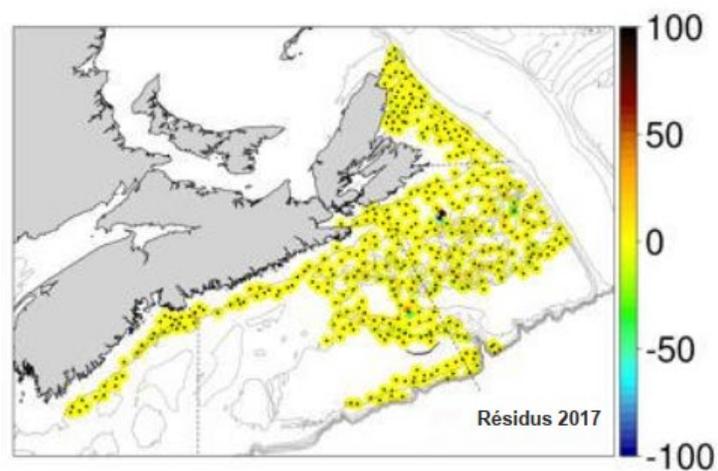
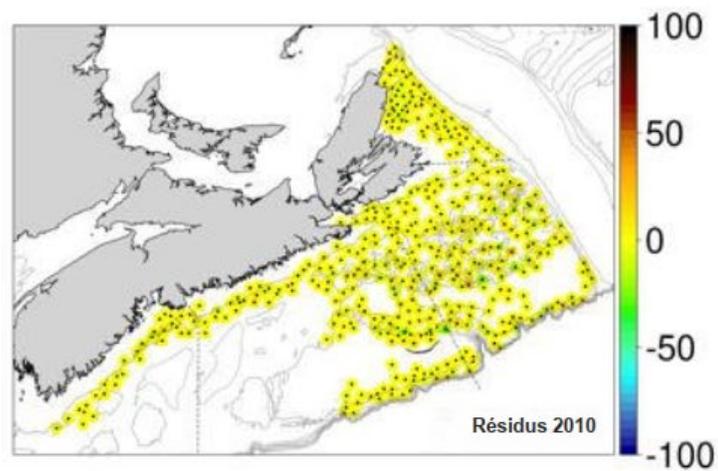
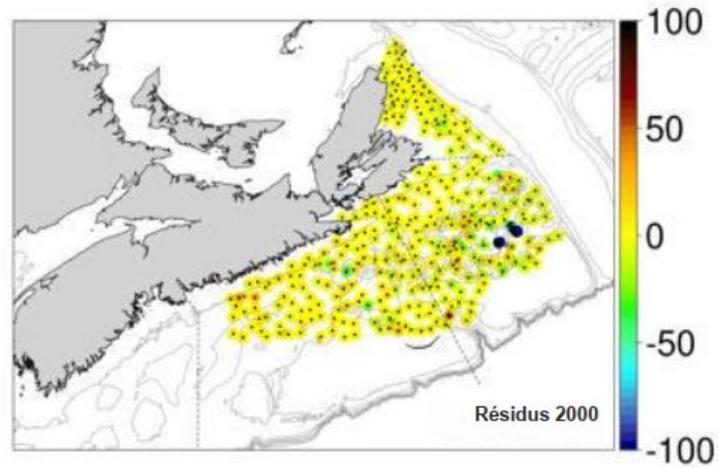


Figure A1.10. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges, sur l'échelle du nombre d'individus par calée—**Simple non séparable**.

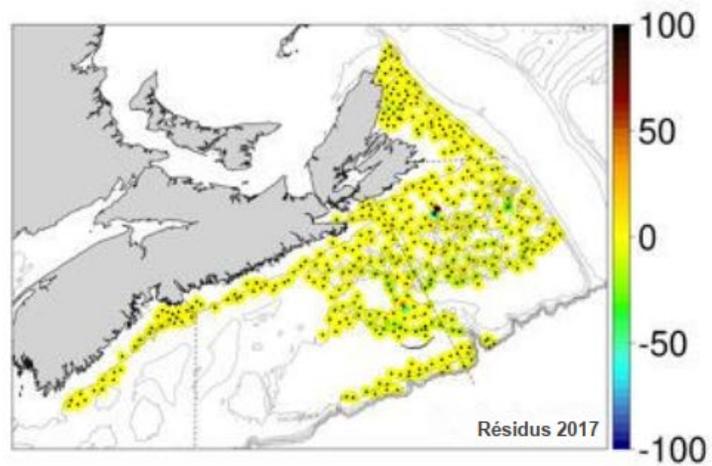
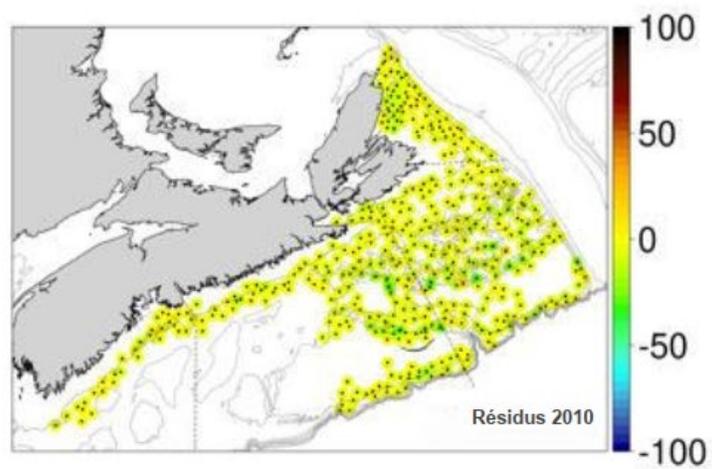
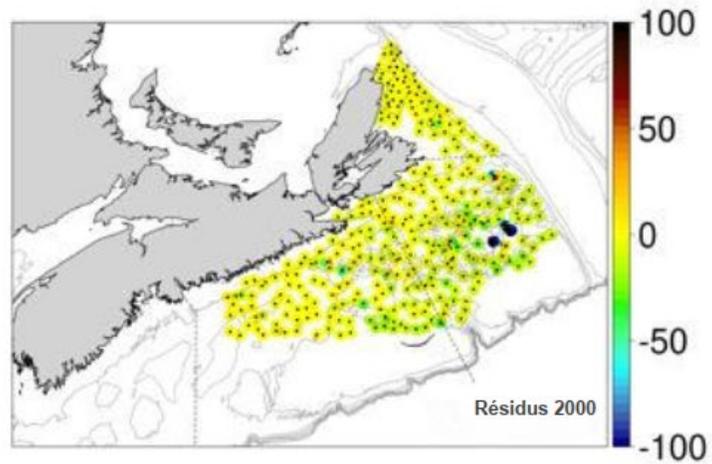


Figure A1.11. Résidus (observés-prévus) des solutions modélisées dans le relevé du crabe des neiges, sur l'échelle du nombre d'individus par calée—**Non séparable**.

ANNEXE 2 : RÉSULTATS DU MODÈLE

A2.1. RÉSULTATS DU MODÈLE « FACTORIEL CROISÉ ». IL S'AGIT DU MLG UTILISANT L'INLA POUR ESTIMER L'ABONDANCE AVEC UNE HYPOTHÈSE DE RÉPARTITION DE POISSON, L'ESPACE ET LE TEMPS ÉTANT TRAITÉS COMME DES FACTEURS. IL N'Y A PAS D'AUTOCORRÉLATION

Formula:

```
inla( formula = totno ~ -1 + year_factor:avid + offset(log(data_offset)),  
      family = poisson)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
year_factor1999:avid1	-0.986	22.483	-54.061	3.115	31.004	15.538	0
...							
year_factor2018:avid190	-1.094	21.889	-52.781	2.906	30.010	15.062	0

Expected number of effective parameters(stdev): 2590.09(0.00)
Number of equivalent replicates : 2.94
Deviance Information Criterion (DIC): NA
Deviance Information Criterion (DIC, saturated): NA
Effective number of parameters: NA
Marginal log-Likelihood: -34659

A2.2. RÉSULTATS DU MODÈLE GLOBAL STMV F() POUR LA TAILLE DES GRAINS DU SUBSTRAT DÉRIVÉS DES MODÈLES ADDITIFS GÉNÉRALISÉS

Family: gaussian

Link function: log

Formula:

```
substrate.grainsize ~ s(b.sdSpatial, k = 3, bs = ts) + s(b.localrange,  
  k = 3, bs = ts) + s(log(z), k = 3, bs = ts) + s(log(dZ),  
  k = 3, bs = ts) + s(log(ddZ), k = 3, bs = ts)
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.9053906	0.0111006	-81.5626	< 2.22e-16

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(b.sdSpatial)	1.99628	2	304.92667	< 2.22e-16
s(b.localrange)	1.99992	2	4461.65359	< 2.22e-16
s(log(z))	1.99666	2	18144.34273	< 2.22e-16
s(log(dZ))	1.98167	2	26.51703	2.2327e-12
s(log(ddZ))	1.97150	2	7.92948	0.00031715

R-sq.(adj) = 0.14 Deviance explained = 13.8%

GCV = 5.6951 Scale est. = 5.695 n = 713021

AIC = 3263839.33

A2.3. RÉSULTATS DU MODÈLE « SIMPLE À EFFETS MIXTES »

Formula:

```
inla(formula = totno ~ 1
+ year_factor
+ f(auid, model="bym2", graph=sppoly@nb, scale.model=TRUE, constr=TRUE, hyper= H$bym2)
+ offset(log(data_offset)),
family="poisson" )
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld	
(Intercept)	6.721	0.088		6.546	6.722	6.894	6.723	0
year_factor2000	0.101	0.068		-0.031	0.101	0.235	0.100	0
...								
year_factor2018	-0.808	0.061		-0.926	-0.808	-0.687	-0.809	0

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	
Precision for auid	0.313	0.059		0.201	0.313	0.428	0.316
Phi for auid	0.836	0.095		0.621	0.849	0.972	0.906

A2.4. RÉSULTATS DU MODÈLE « STATIQUE À EFFETS MIXTES »

Formula:

```
inla(formula = totno ~ 1 + year_factor
+ f(inla.group(z, method="quantile", n=9), model="rw2", scale.model=TRUE, hyper=H$rw2)
+ f(inla.group(substrate.grainsize, method="quantile", n=9 ), model="rw2",
scale.model=TRUE, hyper=H$rw2)
+ f(auid, model="bym2", graph=sppoly@nb, scale.model=TRUE, constr=TRUE, hyper= H$bym2)
+ offset(log(data_offset)),
family="poisson")
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld	
(Intercept)	6.519	0.086		6.349	6.520	6.686	6.521	0
year_factor2000	0.160	0.068		0.027	0.160	0.295	0.159	0
...								
year_factor2018	-0.706	0.062		-0.826	-0.707	-0.584	-0.707	0

Model hyperparameters:

	mean	sd
Prec inla.group(z, method = "quantile", n = 9)	4.449	1.779
Prec inla.group(substrate.grainsize, method = "quantile", n = 9)	0.211	0.056
Prec auid	0.298	0.047
Phi for auid	0.889	0.070

A2.5. RÉSULTATS DE L'ESTIMATION DE LA TEMPÉRATURE AU FOND – CAR POUR CHAQUE ANNÉE

Formula:

```
inla(formula = t ~ 1
      + f( dyri, model=ar1, hyper=H$ar1 )
      + f( inla.group( z, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
      + f( auid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE, constr=TRUE,
          hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group)),
      family = normal )
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	4.52	0.285	3.937	4.522	5.089	4.524	0

Model hyperparameters:

	mean	sd
Precision for the Gaussian observations	0.504	0.005
Precision for dyri	6.302	2.551
Rho for dyri	0.725	0.121
Precision for inla.group(z, method = quantile, n=9)	0.413	0.111
Precision for auid	0.427	0.024
Phi for auid	0.427	0.034
GroupRho for auid	0.841	0.011
Deviance Information Criterion (DIC)	96494.96	
Deviance Information Criterion (DIC, saturated)	28729.33	
Effective number of parameters	1862.94	
Marginal log-Likelihood:	-47560.07	

A2.6. RÉSULTATS DE L'AXE 1 DE LA COMPOSITION DE L'ESPÈCE – CAR POUR CHAQUE ANNÉE

Formula:

```
inla(formula = pca1 ~ 1
      + f( dyri, model=ar1, hyper=H$ar1 )
      + f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
      + f( inla.group( z, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
      + f( inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2, scale.model=TRUE,
          hyper=H$rw2)
      + f( auid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE, constr=TRUE,
          hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group)),
      family = normal)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	0.143	0.015	0.11	0.143	0.173	0.144	0

Model hyperparameters:

	mean	sd
Prec the Gaussian observations	1.91e+02	2.88e+00
Prec dyri	1.03e+03	6.60e+02
Rho for dyri	5.50e-02	2.47e-01
Prec inla.group(t, method = quantile, n=9)	1.70e+04	1.45e+04
Prec inla.group(z, method = quantile, n=9)	1.61e+02	7.50e+01
Prec inla.group(substrate.grainsize, method = quantile, n=9)	7.29e+02	2.72e+03
Prec auid	1.71e+02	1.21e+01
Phi for auid	5.19e-01	3.80e-02
GroupRho for auid	9.19e-01	7.00e-03

Expected number of effective parameters(stdev): 1129.49(30.74)
Number of equivalent replicates : 9.58
Deviance Information Criterion (DIC): -24998.14
Deviance Information Criterion (DIC, saturated): 11953.32
Effective number of parameters: 1132.66
Marginal log-Likelihood: 13670.07

A2.7. RÉSULTATS DE L'AXE 2 DE LA COMPOSITION DE L'ESPÈCE – CAR POUR CHAQUE ANNÉE

Formula:

```
inla(formula = pca2 ~ 1
+ f( dyri, model=ar1, hyper=H$ar1 )
+ f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( z, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2,
scale.model=TRUE, hyper=H$rw2)
+ f( auid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE, constr=TRUE,
hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group)),
family = normal)
```

Fixed effects:

```
      mean    sd 0.025quant 0.5quant 0.975quant mode kld
(Intercept) 0.02 0.021   -0.024    0.02    0.065 0.02  0
```

Model hyperparameters:

	mean	sd
Prec the Gaussian observations	255.139	3.86e+00
Prec dyri	972.573	5.21e+02
Rho for dyri	0.601	2.19e-01
Prec inla.group(t, method = quantile, n=9)	7071.547	1.04e+04
Prec inla.group(z, method = quantile, n=9)	340.448	2.10e+02
Prec inla.group(substrate.grainsize, method = quantile, n=9)	1267.298	1.65e+03
Prec auid	398.917	3.31e+01
Phi for auid	0.538	7.00e-02
GroupRho for auid	0.914	8.00e-03

Expected number of effective parameters(stdev): 921.94(40.80)
Number of equivalent replicates : 11.74
Deviance Information Criterion (DIC): -28335.04
Deviance Information Criterion (DIC, saturated): 11769.79
Effective number of parameters: 928.42
Marginal log-Likelihood: 15491.21

A2.8. RÉSULTATS DU MODÈLE « DYNAMIQUE À EFFETS MIXTES »

Formula:

```
inla(formula = totno ~ 1
+ year_factor
+ f( dyri, model=ar1, hyper=H$ar1 )
+ f(inla.group( z, method=quantile, n=9), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2,
scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca1, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca2, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(auid, model=bym2, graph=sppoly@nb, scale.model=TRUE, constr=TRUE, hyper= H$bym2)
+ offset(log(data_offset)),
family=poisson)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	6.162	0.084	5.996	6.162	6.325	6.163	0
year_factor2000	0.245	0.069	0.111	0.245	0.382	0.244	0

Model hyperparameters:

	mean	sd
Prec inla.group(z, method = quantile, n = 9)	15.418	7.577
Prec inla.group(substrate.grainsize, method = quantile, n = 9)	0.532	0.200
Prec inla.group(t, method = quantile, n = 9)	169.717	111.319
Prec inla.group(pca1, method = quantile, n = 9)	8.211	3.751
Prec inla.group(pca2, method = quantile, n = 9)	20.566	10.102
Prec auid	0.551	0.083
Phi for auid	0.757	0.082

A2.9. RÉSULTATS DU MODÈLE « SÉPARABLE »

Formula:

```
inla(formula = totno ~ 1
+ f( year_factor, model=ar1, hyper=H$ar1 )
+ f( dyri, model=ar1, hyper=H$ar1 )
+ f(inla.group( z, method=quantile, n=9), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2,
scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca1, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca2, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(aid, model=bym2, graph=sppoly@nb, scale.model=TRUE, constr=TRUE, hyper= H$bym2)
+ offset(log(data_offset)),
family=poisson)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	5.955	0.149	5.663	5.955	6.248	5.955	0.001

Model hyperparameters:

	mean	sd
Prec year_factor	8.383	2.925
Rho for year_factor	0.862	0.051
Prec inla.group(z, method = quantile, n = 9)	16.377	12.415
Prec inla.group(substrate.grainsize, method = quantile, n = 9)	0.522	0.194
Prec inla.group(t, method = quantile, n = 9)	173.327	113.614
Prec inla.group(pca1, method = quantile, n = 9)	8.143	3.786
Prec inla.group(pca2, method = quantile, n = 9)	19.974	12.119
Prec aid	0.491	0.098
Phi for aid	0.849	0.092

A2.10. RÉSULTATS DU MODÈLE « SIMPLE NON SÉPARABLE »

Formula:

```
inla(formula = totno ~ 1
+ f( aid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE,
constr=TRUE, hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group))
+ offset(log(data_offset)),
family=poisson)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	5.838	0.083	5.672	5.839	5.997	5.841	0

Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
Precision for aid	0.283	0.024	0.238	0.282	0.334	0.280
Phi for aid	0.622	0.051	0.525	0.621	0.723	0.615
GroupRho for aid	0.886	0.009	0.866	0.887	0.904	0.888

A2.11. RÉSULTATS DU MODÈLE « ESPACE|TEMPS NON SÉPARABLE »

Formula:

```
inla(formula = totno ~ 1
+ f( dyri, model=ar1, hyper=H$ar1 )
+ f(inla.group( z, method=quantile, n=9), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2,
scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca1, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca2, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( auid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE,
constr=TRUE, hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group))
+ offset(log(data_offset)),
family=poisson,
data = M,
verbose = TRUE,
control.compute = list(dic = TRUE, config = TRUE),
control.predictor = list(compute = FALSE, link = 1),
control.inla = list(cmin = 0, h = 1e-06, tolerance = 1e-12),
control.results = list(return.marginals.random = TRUE, return.marginals.predictor = TRUE),
control.fixed = H$fixed
)
```

Fixed effects:Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	5.547	0.11	5.321	5.55	5.759	5.554	0

Model hyperparameters:

	mean	sd
Precision for dyri	28.692	16.806
Rho for dyri	-0.012	0.143
Precision for inla.group(t, method = quantile, n = 9)	256.897	330.921
Precision for inla.group(z, method = quantile, n = 9)	8.575	3.873
Precision for inla.group(substrate.grainsize, method = quantile, n = 9)	1.460	0.859
Precision for inla.group(pca1, method = quantile, n = 9)	8.077	3.837
Precision for inla.group(pca2, method = quantile, n = 9)	23.812	11.599
Precision for auid	0.435	0.035
Phi for auid	0.661	0.060
GroupRho for auid	0.805	0.015

A2.12. RÉSULTATS DU MODÈLE « HABITAT ESPACE|TEMPS NON SÉPARABLE » POUR LE PETIT CRABE IMMATURE (LARGEUR DE CARAPACE < 50 MM)

Formula:

```
inla(formula = pa ~ 1
+ f( dyri, model=ar1, hyper=H$ar1 )
+ f(inla.group( z, method=quantile, n=9), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f(inla.group( substrate.grainsize, method=quantile, n=9 ), model=rw2,
scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( t, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca1, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( inla.group( pca2, method=quantile, n=9 ), model=rw2, scale.model=TRUE, hyper=H$rw2)
+ f( auid, model=bym2, graph=sppoly@nb, group=year_factor, scale.model=TRUE,
constr=TRUE, hyper=H$bym2, control.group=list(model=ar1, hyper=H$ar1_group)),
family=binomial)
```

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	-0.384	0.192	-0.792	-0.376	-0.017	-0.364	0

Model hyperparameters:

	mean	sd
Precision for dyri	13.338	13.600
Rho for dyri	0.024	0.211
Precision for inla.group(t, method = quantile, n=9)	142.645	387.404
Precision for inla.group(z, method = quantile, n=9)	2.163	1.742
Precision for inla.group(substrate.grainsize, method = quantile, n=9)	16.309	53.300
Precision for inla.group(pca1, method = quantile, n=9)	30.908	37.211
Precision for inla.group(pca2, method = quantile, n=9)	645.305	6052.288
Precision for auid	0.426	0.057
Phi for auid	0.974	0.026
GroupRho for auid	0.921	0.014
Expected number of effective parameters(stdev):	408.60	(22.44)
Number of equivalent replicates :	18.61	
Deviance Information Criterion (DIC)	6872.43	
Deviance Information Criterion (DIC, saturated)	6872.43	
Effective number of parameters	389.55	
Marginal log-Likelihood:	-1868.98	