



Pêches et Océans
Canada

Fisheries and Oceans
Canada

Sciences des écosystèmes
et des océans

Ecosystems and
Oceans Science

Secrétariat canadien de consultation scientifique (SCCS)

Document de recherche 2020/004

Région du Pacifique

Élaboration d'un cadre de modélisation de la répartition des espèces et son application à douze espèces sur la côte canadienne du Pacifique

Nephtin¹, J., E.J. Gregr², C. St. Germain¹, C. Fields¹, J.L. Finney¹

¹ Pêches et Océans Canada
Station biologique du Pacifique
3190, chemin Hammond Bay
Nanaimo (Colombie-Britannique) V9T 6N7

² SciTech Environmental Consulting
2136, rue Napier
Vancouver (Colombie-Britannique) V5L 2N9

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 de consultation scientifique
200, rue Kent
Ottawa (Ontario) K1A 0E6

[http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca](http://www.dfo-mpo.gc.ca/csas-sccs/csas-sccs@dfo-mpo.gc.ca)



© Sa Majesté la Reine du chef du Canada, 2020
ISSN 2292-4272

La présente publication doit être citée comme suit :

Nepin, J., Gregr, E.J., St. Germain, C., Fields, C., et Finney, J.L. 2020. Élaboration d'un cadre de modélisation de la répartition des espèces et son application à douze espèces sur la côte canadienne du Pacifique. Secr. can. de consult. sci. du MPO. Doc. de rech. 2020/004. xiv + 118 p.

Also available in English :

Nepin, J., Gregr, E.J., St. Germain, C., Fields, C., and Finney, J.L. 2020. Development of a Species Distribution Modelling Framework and its Application to Twelve Species on Canada's Pacific Coast. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/004. xii + 107 p.

TABLE DES MATIÈRES

RÉSUMÉ	xiv
1 INTRODUCTION.....	1
1.1 APERÇU DE LA MODÉLISATION DE LA RÉPARTITION DES ESPÈCES	1
1.1.1 Le modèle écologique.....	5
1.1.2 Le modèle de données	5
1.1.3 Le modèle statistique	6
1.1.4 Intégration des connaissances écologiques locales	6
2 SÉLECTION DES DONNÉES	7
2.1 DONNÉES SUR L'OCCURRENCE	7
2.1.1 Taille de l'échantillon.....	8
2.1.2 Étendues	8
2.1.3 Biais d'échantillonnage	9
2.1.4 Précision.....	9
2.2 DONNÉES PRÉDICTIVES	9
2.2.1 Étendues et résolution	10
2.2.2 Précision.....	11
2.2.3 Mise à l'échelle	11
2.2.4 Prédicteurs spatiaux	12
3 UN CADRE DE MODÉLISATION DE LA RÉPARTITION DES ESPÈCES	12
3.1 PRÉPARATION DES DONNÉES	14
3.2 VALIDATION CROISÉE.....	15
3.3 AJUSTEMENT DU MODÈLE	17
3.4 ÉVALUATION DU MODÈLE	18
3.5 PRÉDICTION.....	21
3.6 INCERTITUDE.....	23
3.7 INTERPRÉTATION DES MODÈLES	27
3.8 LE RÔLE DE LA STRUCTURE SPATIALE.....	29
4 UNE APPLICATION DU CADRE DE MODÉLISATION DE LA RÉPARTITION DES ESPÈCES.....	32
4.1 INTRODUCTION	32
4.2 MÉTHODES	33
4.2.1 Zones d'étude	34
4.2.2 Prédicteurs environnementaux	35
4.2.3 Observations des espèces.....	39
4.2.4 Préparation des données	43
4.2.5 Modèles d'indice de qualité de l'habitat.....	43
4.2.6 Modèles linéaires généralisés	44
4.2.7 Arbres de régression stimulée	45

4.2.8	Modèles d'ensemble	46
4.2.9	Évaluation du modèle	46
4.3	RÉSULTATS ET DISCUSSION	46
4.3.1	Performance du modèle par espèce	46
4.3.2	Comparaisons des prédictions des modèles entre les méthodes	47
4.3.3	Incertitude des modèles	52
4.3.4	Comparaison des statistiques sur la performance	54
4.4	CONCLUSIONS	56
5	RECOMMANDATIONS	56
6	GLOSSAIRE	58
7	REMERCIEMENTS	60
8	RÉFÉRENCES	61
ANNEXE A. PRÉDICTIONS DE LA RÉPARTITION DE DOUZE ESPÈCES À PARTIR DE L'APPLICATION DU CADRE		70
	Panope du Pacifique	71
	Oursin rouge	74
	Varech pterygophora	77
	Zostère marine	80
	Crabe dormeur	83
	Sébaste à dos épineux	86
	Sébaste aux yeux jaunes	89
	Étoile ocrée	92
	Complexe de moules bleues	93
	Palourde du Pacifique	94
	Pennatule orangée	96
ANNEXE B. DESCRIPTIONS DE L'INDICE DE QUALITÉ DE L'HABITAT		97
	Ormeau nordique	98
	Panope du Pacifique	100
	Oursin rouge	101
	Varech pterygophora	103
	Zostère marine	104
	Crabe dormeur	105
	Sébaste à dos épineux	107
	Sébaste aux yeux jaunes	109
	Étoile ocrée	110
	Complexe de moules bleues	111
	Palourde du Pacifique	113
	Pennatule orangée	114
	RÉFÉRENCES ANNEXE B	115

LISTE DES TABLEAUX

Tableau 4.1. Sources des variables prédictives environnementales utilisées pour modéliser la qualité de l'habitat de 12 espèces à deux résolutions spatiales (20 m pour les eaux côtières et 100 m pour le plateau). Voir pour la description des unités pour chaque couche de prédicteurs à la section 4.2.2.....	37
Tableau 4.2. Sommaire des données sur l'occurrence choisies pour modéliser les douze espèces dans cette étude. Une régression linéaire généralisée et des arbres à régression stimulée ont été utilisés pour construire des modèles pour les huit espèces pour lesquelles on n'a pas jugé que les données étaient insuffisantes (DI). Les relevés indiqués avec des données sources (ABL = ormeau; RSU = oursin rouge; BHM = cartographie de l'habitat benthique; cuke = holothurie; HBLL = palangre sur fond dur; IPHC = Commission internationale du flétan du Pacifique) représentent les relevés scientifiques réalisés pour les évaluations ou la surveillance des stocks menées par la région du Pacifique du MPO et ses partenaires industriels.....	42
Tableau 4.3. Paramètres initiaux pour les modèles d'arbres de régression stimulée.....	45

LISTE DES FIGURES

Figure 1.1. Aperçu du processus complet de modélisation, de l'élaboration du contexte du modèle (contextualisation) à l'évaluation des prédictions du modèle. Le cadre élaboré ici pour la modélisation de la répartition des espèces comprend une série d'étapes prescrites qui automatisent les pratiques exemplaires. Les modèles linéaires généralisés (MLG) sont un exemple de méthode de modélisation axée sur les données, tandis que les modèles d'indice de qualité de l'habitat (IQH) sont un exemple d'approche de modélisation d'enveloppe fondée sur les connaissances.....	4
Figure 3.1. Diagramme du flux du cadre mis en œuvre avec des scripts spécialement conçus à cet effet. Le flux comprend le traitement des observations (données dépendantes) et des couches de prédicteurs, ainsi que l'ajustement, l'évaluation et la prédiction des modèles selon une approche de validation croisée. L'approche d'ajustement du modèle varie selon le choix de la méthode de modélisation. Les modèles linéaires généralisés (MLG) et les modèles à arbre de régression stimulée (ARS) sont deux exemples de méthodes d'ajustement des modèles. Les cases bleues pleines représentent les extraits du processus de modélisation.	13
Figure 3.2. Distribution des observations de présence et d'absence d'une espèce dans une zone d'étude.	15
Figure 3.3. Orientation spatiale de la validation croisée par blocs spatiaux à cinq plis d'une espèce dans une zone d'étude. Chaque pli serait utilisé une fois comme données d'essai pour évaluer les modèles élaborés avec les données d'apprentissage constituées des plis restants.	16
Figure 3.4. Exemples de relations possibles entre les variables prédictives environnementales et la qualité de l'habitat utilisées pour élaborer des modèles d'indice de qualité de l'habitat.....	17
Figure 3.5. Courbe de fonction d'efficacité du récepteur (ROC). Les lignes grises représentent les courbes ROC des modèles à validation croisée à cinq plis et la ligne noire représente la moyenne des courbes ROC. Les courbes sont établies à partir des données d'essai. AUC désigne l'aire sous la courbe ROC. Le point noir indique le taux de faux positifs et de vrais positifs au seuil optimisé.	20
Figure 3.6. Courbes représentant les mesures de l'exactitude d'un modèle dépendant du seuil sur une gamme de seuils. Les courbes sont construites à partir des données d'essai. Les rubans gris représentent un écart-type autour de la moyenne calculée à partir de modèles à VC à cinq plis. TSS désigne la « true skill statistic ».	20
Figure 3.7. Représentation visuelle d'une matrice de contingence montrant la densité des prédictions de positifs vrais et faux au-dessus du seuil et des prédictions de négatifs vrais et faux au-dessous du seuil. Le seuil est représenté par une ligne pointillée grise. Les polygones noirs sont des estimations de la densité du noyau qui illustrent la densité des prédictions correspondant aux observations de présence et d'absence.....	21
Figure 3.8. Diagramme du flux du modèle d'ensemble montrant comment les prédictions issues de différentes méthodes de modélisation peuvent être combinées en un ensemble dans la structure de validation croisée. Chaque modèle, y compris l'ensemble, est validé avec un jeu identique de données d'essai pour chaque procédure de validation croisée. Les modèles linéaires généralisés (MLG), l'arbre de régression simulée (ARS), la forêt d'arbres décisionnels (FA) et les modèles additifs généralisés (MAG) sont quatre exemples de méthodes d'ajustement des modèles.....	22
Figure 3.9. Prédictions de la probabilité d'occurrence à partir d'un modèle d'enveloppe fondé sur les connaissances (p. ex. IQH, à gauche) et d'un modèle d'ensemble axé sur les données	

(à droite). La prédiction d'ensemble est la moyenne des prédictions des modèles à validation croisée à cinq plis élaborés à partir de méthodes multiples (p. ex. MLG et ARS).....	23
Figure 3.10. Zones extrapolées où les prédicteurs se situaient à l'extérieur de la plage des valeurs utilisées pour l'ajustement du modèle (à gauche). Prédictions de la probabilité d'occurrence moyenne à partir de modèles à validation croisée à cinq plis, en excluant les zones extrapolées dans l'espace des prédicteurs (à droite). La ligne noire représente le plus petit polygone convexe autour des observations.....	24
Figure 3.11. Différence entre les prédictions d'un modèle d'enveloppe fondé sur les connaissances et d'un modèle d'ensemble fondé sur les données. Les valeurs positives représentent les zones où le modèle d'ensemble prédisait une plus grande probabilité d'occurrence que le modèle d'enveloppe.	25
Figure 3.12. Écart-type des prédictions de la probabilité d'occurrence à partir de modèles à validation croisée à cinq plis.....	26
Figure 3.13. Moyenne des résiduels du modèle (agrégés à 10 km) à partir de modèles à validation croisée à cinq plis.....	27
Figure 3.14. Influence relative moyenne des prédicteurs des modèles à validation croisée à cinq plis. Les barres d'erreur représentent l'influence relative minimale et maximale des cinq modèles à VC.	28
Figure 3.15. Effets marginaux moyens, minimaux et maximaux entre chaque variable prédictive et la réponse prédite des modèles à validation croisée à cinq plis. Les effets marginaux ont été calculés à partir de modèles linéaires généralisés. La ligne pleine noire représente la moyenne, et la zone grisée représente les effets marginaux minimaux et maximaux des cinq modèles. Les effets marginaux sont présentés pour toutes les couches de prédicteurs environnementaux. La ligne pointillée indique une extrapolation dans l'espace des prédicteurs où la variable prédictive se situe en dehors de la plage des valeurs utilisées pour l'ajustement du modèle, ce qui indique qu'il n'y a eu aucune observation de l'espèce dans cet espace. Les marques le long de l'axe des abscisses indiquent la densité des observations dans la plage des valeurs prédictives.....	29
Figure 3.16. Variogramme des observations de la présence-absence des espèces et des résiduels du modèle.....	30
Figure 3.17. Distribution d'un prédicteur autocovariable spatial qui représente la somme pondérée par la distance des observations proches.....	31
Figure 3.18. Influence relative moyenne des prédicteurs des modèles à validation croisée à cinq plis, y compris l'autocovariable. Les barres d'erreur représentent l'influence relative minimale et maximale des cinq modèles.	32
Figure 4.1. Les deux zones d'étude utilisées dans cette analyse. La zone d'étude des eaux côtières du nord de la côte centrale (résolution de 20 m) couvre environ 5 620 km ² . La zone d'étude du plateau (résolution de 100 m) comprend les biorégions marines des plateaux du nord et du sud et couvre environ 127 840 km ²	34
Figure 4.2. Performance des huit modèles de répartition pour les espèces pour lesquelles les données sur l'occurrence sont suffisantes. L'aire moyenne sous la courbe de fonction d'efficacité du récepteur (AUC) est basée sur des essais de validation croisée par blocs spatiaux à cinq plis et est indiquée pour chacun des quatre types de modèles : indice de qualité de l'habitat (IQH), modèle linéaire généralisé (MLG), arbre à régression stimulée (ARS) et ensemble. Les barres d'erreur représentent un écart-type. La prévalence et la taille de l'échantillon sont indiquées dans l'angle supérieur gauche. Les barres non représentées ont une	

valeur AUC inférieure à 0,5. Les modèles IQH et d'ensemble n'ont pas été évalués à l'aide de données d'apprentissage.49

Figure 4.3. Influence relative des prédicteurs et effets marginaux des huit prédicteurs environnementaux les plus influents de la répartition de l'ormeau d'après les modèles IQH, MLG et ARS. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne et les lignes d'erreur montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Sur les courbes des effets marginaux, les lignes pleines noires représentent la moyenne et la zone grisée représente les effets marginaux minimaux et maximaux des cinq modèles à VC à cinq plis. Pour représenter le substrat, on a utilisé une couche de substrat catégoriel pour le modèle IQH et un indice continu de roccosità pour les modèles MLG et ARS.50

Figure 4.4. Diagrammes montrant la distribution des paramètres de performance des modèles à partir des modèles d'indice de qualité de l'habitat (IQH), des modèles linéaires généralisés (MLG), des modèles d'arbre à régression stimulée (ARS) et des modèles d'ensemble parmi tous les modèles d'espèces mis à l'essai. AUC désigne l'aire sous la courbe de fonction d'efficacité du récepteur, et TSS désigne la « true skill statistic »51

Figure 4.5. Prédictions de la probabilité d'occurrence des espèces dans la zone d'étude du plateau d'après le modèle d'indice de qualité de l'habitat (IQH), le modèle linéaire généralisé (MLG), le modèle de l'arbre de régression stimulée (ARS) et le modèle d'ensemble du sébaste à dos épineux. À l'exception de la prédiction du IQH, la probabilité d'occurrence est la prédiction moyenne des modèles à validation croisée à cinq plis. Les prédictions MLG, ARS et d'ensemble sont tronquées pour exclure les zones d'extrapolation dans l'espace des prédicteurs.52

Figure 4.6. Distribution de l'incertitude (en tant qu'écart-type moyen entre les prédictions des espèces) pour trois méthodes de modélisation. L'écart-type moyen des prédictions est basé sur les prédictions des modèles à validation croisée à cinq plis pour toutes les espèces. Les méthodes comprenaient le modèle linéaire généralisé (MLG), le modèle d'arbre de régression stimulée (ARS) et le modèle d'ensemble. L'écart-type des modèles d'ensemble incorpore la variance entre les méthodes MLG et ARS, ainsi que les différents modèles à VC.53

Figure 4.7. Incertitude (en tant qu'écart-type moyen par rapport aux prédictions du modèle d'ensemble) pour les huit espèces pour lesquelles les données d'occurrence sont suffisantes, de la plus élevée à la plus faible.54

Figure 4.8. Corrélation entre les prédicteurs fondés sur un seuil, exactitude, kappa et TSS (statistique de la qualité vraie) et l'aire sous la courbe de fonction d'efficacité du récepteur (AUC). Toutes les espèces et méthodes de modélisation ont été utilisées pour la comparaison.55

Figure 4.9. Performance moyenne pour le kappa et la TSS (« true skill statistic ») à partir d'une validation croisée à cinq plis, pour des espèces à prévalence élevée et faible. Les paramètres de validation sont présentés séparément pour les modèles linéaires généralisés (MLG) et d'arbre de régression stimulée (ARS). La prévalence est indiquée dans l'angle supérieur gauche pour chaque espèce. Les barres d'erreur représentent un écart-type.55

Figure A.1. Observations de la présence et de l'absence de la panope dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.2.71

Figure A.2. Prédictions de la répartition de la panope et incertitude correspondante. Prédictions de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du

modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.....72

Figure A.3. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS de la panope. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice continu du niveau de sable pour les modèles MLG et ARS.73

Figure A.4. Observations de la présence et de l'absence de l'oursin rouge dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.5.....74

Figure A.5. Prédications de la répartition de l'oursin rouge et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée à la figure A.4 par la case en pointillés.75

Figure A.6. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS de l'oursin rouge. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.76

Figure A.7. Observations de la présence et de l'absence du varech *Pterygophora* dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée sur la figure A.8. 77

Figure A.8. Prédications de la répartition du varech *Pterygophora* et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée sur la figure A.7 par la case en pointillés.78

Figure A.9. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du varech *Pterygophora*. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens

par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.79

Figure A.10. Observations de la présence et de l'absence de la zostère marine dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.11. .80

Figure A.11. Prédications de la répartition de la zostère marine et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée sur la figure A.10 par la case en pointillés.81

Figure A.12. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS de la zostère marine. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice continu du niveau de sable pour les modèles MLG et ARS.82

Figure A.13. Observations de la présence et de l'absence du crabe dormeur dans la zone d'étude du plateau.83

Figure A.14. Prédications de la répartition du crabe dormeur et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.84

Figure A.15. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du crabe dormeur. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice continu du niveau de sable pour les modèles MLG et ARS.85

Figure A.16 Observations de la présence et de l'absence du sébaste à dos épineux dans la zone d'étude du plateau.86

Figure A.17. Prédications de la répartition du sébaste à dos épineux et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la

différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.87

Figure A.18. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédictors environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du sébaste à dos épineux. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.88

Figure A.19 Observations de la présence et de l'absence du sébaste aux yeux jaunes dans la zone d'étude du plateau.89

Figure A.20. Prédiction de la répartition du sébaste aux yeux jaunes et incertitude correspondante. Prédiction de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.90

Figure A.21. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédictors environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du sébaste aux yeux jaunes. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.91

Figure A.22. Prédiction de la répartition de l'étoile ocrée à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée sur la figure A.1 par la case en pointillés.92

Figure A.23. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de l'étoile ocrée.92

Figure A.24. Prédiction de la répartition de la moule bleue à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.93

Figure A.25. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la moule bleue.93

Figure A.26. Prédiction de la répartition de la palourde du Pacifique à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.94

Figure A.27. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la palourde du Pacifique.95

Figure A.28. Prédiction de la répartition de la pennatule orangée à partir d'un modèle d'indice de qualité de l'habitat.96

Figure A.29. Influence relative des prédictors (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la pennatule orangée.96

RÉSUMÉ

Les modèles de répartition des espèces (MRE) sont devenus de plus en plus précieux comme outil de gestion et de conservation des ressources et des lieux marins. Cependant, malgré leur utilité et leur prévalence, les aspects essentiels de l'élaboration de MRE et de l'évaluation de l'incertitude sont régulièrement négligés. Des pratiques exemplaires sont donc justifiées, et leur application est de plus en plus nécessaire. Le présent cadre, mis en œuvre à l'aide de scripts spécialement conçus à cet effet et rédigés dans le langage de programmation statistique R, a été préparé à la fois comme un outil et comme un ensemble de lignes directrices et de méthodes en vue d'élaborer des MRE cohérents, interprétables et défendables pour appuyer la contribution du MPO aux politiques océaniques du Canada. Des MRE ont été élaborés pour douze espèces benthiques afin d'illustrer l'application du cadre et de guider la planification des interventions d'urgence en cas de déversement d'hydrocarbures dans le cadre du Plan d'intervention régional pour la biorégion du plateau du nord. Trois méthodes d'établissement de modèles de plus en plus complexes ont été appliquées à l'aide d'une série des meilleurs prédicteurs environnementaux disponibles. Des modèles d'enveloppe fondés sur les connaissances ont été produits pour toutes les espèces, et l'accent a été mis sur les espèces sur lesquelles les données étaient insuffisantes. Ces modèles d'enveloppe ont guidé l'élaboration de modèles subséquents axés sur les données et on peut les utiliser pour aider à évaluer l'incertitude des prévisions des modèles. Ils donnent également l'occasion de faire participer des experts des espèces au processus. Des modèles linéaires généralisés (MLG) et des modèles à arbre de régression stimulée axés sur les données ont été produits, ainsi qu'un modèle d'ensemble correspondant, pour les huit espèces pour lesquelles les données d'observation étaient adéquates. Des prédictions de la plus haute qualité ont été produites pour les espèces pour lesquelles la taille des échantillons était grande, et les données d'observation étaient bien réparties dans la zone d'étude. Tous les modèles d'ensemble ont obtenu des résultats modérément bons ($AUC > 0,7$) lorsqu'ils ont été évalués selon une approche de validation croisée par blocs spatiaux. Treize recommandations ont été formulées dans le cadre de l'élaboration du cadre et de son application. Elles donnent une orientation sur l'application des méthodes d'élaboration de MRE en ce qui concerne la sélection et la préparation des données, l'élaboration et l'évaluation des modèles, et mettent en évidence les moyens d'améliorer les résultats de la modélisation en appliquant des pratiques exemplaires dans la mesure du possible. L'élaboration future de MRE pour d'autres espèces du Pacifique canadien sera grandement facilitée par l'ensemble de prédicteurs, de méthodes et d'outils d'évaluation communs réunis ici.

1 INTRODUCTION

Il est de plus en plus important de comprendre la répartition possible des espèces marines et de leurs habitats pour prendre des décisions éclairées sur leur gestion et leur conservation. Comme peu d'espèces marines sont inventoriées, les modèles de répartition des espèces (MRE) sont devenus une approche courante pour estimer la répartition des espèces prisées. Les MRE sont définis ici comme des modèles qui établissent un lien entre les données sur l'occurrence et les conditions environnementales dans les lieux d'occurrence connus (Elith et Graham, 2009). En prédisant la répartition d'une espèce à partir des corrélations entre les observations et les prédicteurs environnementaux, les MRE fournissent un moyen rapide et rentable de prévoir les secteurs où les espèces d'intérêt sont susceptibles de se trouver. Le fait de comprendre la répartition potentielle des espèces peut éclairer diverses activités de gestion, y compris la planification spatiale marine, l'évaluation des répercussions de la pêche et la lutte contre les espèces envahissantes. Les nouvelles applications comprennent les changements dans la répartition des espèces en réaction aux changements climatiques, les répercussions de la pêche commerciale, l'évaluation des stocks, la planification régionale des pêches et les interventions d'urgence en cas de pollution. Les MRE peuvent également relever des lacunes dans les connaissances écologiques, ce qui aide à cibler les futurs efforts de relevé et de recherche.

Cependant, l'élaboration des MRE s'accompagne de plusieurs défis. Citons l'accessibilité croissante de méthodes statistiques et données environnementales perfectionnées, la variété des données sur la présence des espèces et la diversité des méthodes d'échantillonnage. De nombreuses considérations liées aux objectifs du modèle, à la préparation des données, à la sélection des variables et aux méthodes analytiques influent également sur le choix des méthodes appropriées et sur l'interprétation du modèle. Ces considérations compliquent l'établissement de la meilleure méthode pour une application particulière (GREGG et CHAN, 2014; YATES *et al.*, 2018).

Le présent document vise à normaliser et à faciliter le processus d'élaboration des MRE en présentant un cadre pour guider la préparation des données, l'ajustement des modèles, l'évaluation des modèles et l'interprétation des résultats et des incertitudes. Le cadre cible les MRE construits à partir de données sur l'occurrence des espèces (présence-absence). Les estimations de l'abondance ne sont pas prises en compte. Le cadre s'inspire des pratiques exemplaires actuelles, et son application est illustrée par un ensemble diversifié d'espèces benthiques de la côte canadienne du Pacifique. Il est mis en œuvre à l'aide de scripts spécialement conçus à cet effet et rédigés dans le langage de programmation statistique R. Pour réduire le besoin de définir les termes dans le document, un glossaire de la terminologie pertinente est inclus.

1.1 APERÇU DE LA MODÉLISATION DE LA RÉPARTITION DES ESPÈCES

Au fil des ans, les travaux visant à comprendre comment les espèces sont réparties dans l'espace et le temps se sont déroulés individuellement dans diverses disciplines, et divers termes ont été utilisés (Hirzel et Le Lay, 2008). Bien que le cadre cible les MRE axés sur les données, il tient également compte de l'utilité des modèles d'enveloppe, qui fournissent un moyen de délimiter la répartition des espèces à l'aide de données sur la présence seulement, ou simplement de contraintes environnementales hypothétiques. Ils servent ici à décrire la compréhension écologique actuelle de la façon dont les espèces sont liées à leur environnement et à évaluer la mesure dans laquelle cette compréhension écologique se reflète dans les MRE axés sur les données.

Il est généralement admis que la répartition des espèces dépend des interactions biotiques et des obstacles à la dispersion, en plus de facteurs environnementaux (Soberón et Peterson, 2005). Toutefois, la documentation demeure équivoque quant à la question de savoir si les prédictions fondées sur les relations entre les observations des espèces et l'environnement ne reflètent qu'un habitat convenable ou peuvent être étendues à la répartition des espèces. Par exemple, Guisan *et al.* (2002) soutiennent que tous ces modèles sont des MRE parce que les observations des espèces sont nécessairement limitées par ces trois facteurs. D'autres (p. ex. Araújo et Peterson, 2012) ont soutenu que, malgré les contraintes qui pèsent sur les données d'observation, des processus importants ne sont toujours pas pris en compte si l'on n'utilise que des prédicteurs environnementaux. Des travaux récents (Dallas et Hastings, 2018) ont révélé que les hypothèses sur l'abondance et la qualité qui sous-tendent les MRE semblent dépendre de l'étendue et de la résolution de la zone d'étude. Ainsi, il est probable que, en réalité, les modèles fondés sur les corrélations environnementales se situent quelque part dans un continuum entre la qualité de l'habitat et la répartition des espèces, selon l'écologie de l'espèce, ses antécédents d'exploitation, la qualité des données disponibles et l'échelle de l'analyse.

Avant de construire un MRE, il est utile de se pencher sur le contexte global du modèle souhaité. Le contexte du modèle inclut le besoin de gestion, ce que l'on comprend de l'espèce et de son habitat, et les données disponibles (figure 1.1). Une bonne compréhension du contexte du modèle, en plus de la prise en compte des hypothèses requises à chaque étape du processus d'élaboration du modèle, aidera à éclairer les deux méthodes de modélisation les plus appropriées et à faciliter l'interprétation des résultats des modèles.

La pratique exemplaire comprend la définition d'un modèle écologique décrivant le contexte de l'analyse, un modèle de données décrivant les données sur l'occurrence et les prédicteurs, et un modèle statistique reliant les deux. Comme le décrit Austin (2002), le modèle écologique comprend les connaissances écologiques et l'objectif du modèle, qui devrait refléter le besoin en matière de gestion. Le modèle de données comprend la façon dont les données sur l'occurrence et les prédicteurs ont été recueillies, mesurées et organisées. Le modèle statistique décrit les relations entre les données sur l'occurrence et les prédicteurs. La pratique de modélisation acceptée commence par le modèle écologique, ce qui facilite l'élaboration du modèle de données. Ensemble, ces deux modèles éclairent ensuite les étapes de la préparation des données (figure 1.1), et donc la sélection, l'application et l'interprétation du modèle statistique. Par exemple, comprendre comment les données sur l'occurrence ont été échantillonnées peut limiter le stade biologique modélisé, ce qui a des répercussions sur les données utilisées, et ce qui peut être déduit à partir des prédictions du modèle.

Les MRE peuvent varier considérablement en matière de complexité, mais ils sont généralement élaborés à l'aide d'une combinaison de compréhension écologique et d'analyse de corrélation. Les modèles d'enveloppe basés entièrement sur la compréhension écologique peuvent fournir des renseignements utiles à de grandes étendues spatiales (p. ex. Kaschner *et al.*, 2006), tandis que les MRE plus complexes qui reposent principalement sur des corrélations peuvent fournir des prévisions fiables sur de courtes périodes à des échelles locales (p. ex. Maxwell *et al.*, 2015). Le choix des modèles statistiques appropriés dépendra de ce que l'on sait de l'écologie de l'espèce, de l'objectif de gestion et de la disponibilité et de l'applicabilité des données sur l'occurrence et les prédicteurs.

De manière générale, on a supposé que les prédictions des modèles seront plus transférables (c.-à-d. fourniront des prédictions fiables dans des endroits ou à des moments au-delà de la zone d'étude) si les modèles sont moins complexes (Levins, 1966; Yates *et al.*, 2018). Malgré l'existence de preuves à l'appui de ce point de vue (p. ex. Gregr *et al.*, 2018; Randin *et al.*, 2006; Tuanmu *et al.*, 2011), d'autres ont soutenu que ce n'est pas tant la complexité qui pose problème, mais plutôt les relations adaptées à des données biaisées (Evans *et al.*, 2013;

García-Callejas et Araújo, 2016). Néanmoins, on s'entend pour dire que les modèles qui mettent l'accent sur les processus écologiques se prêteront à un meilleur transfert des points de vue du moment et de l'endroit (p. ex. Bell et Schlaepfer, 2016; Yates *et al.*, 2018). Lorsque les objectifs du modèle exigent une transférabilité (p. ex. une expansion ou un déplacement potentiel de l'aire de répartition), il faut essayer de le valider en recueillant des données supplémentaires à différents moments ou endroits (Figure 1.1.). L'évaluation d'un modèle à l'aide de données recueillies de façon indépendante et bien échantillonnées est considérée comme une pratique exemplaire pour évaluer la transférabilité d'un modèle (Araújo *et al.*, 2019).

La modélisation des espèces et de leurs aires de répartition a beaucoup évolué depuis ses origines dans l'écologie des systèmes, alors qu'on élaborait les premières descriptions mathématiques des systèmes écologiques (p. ex. MacArthur, 1955; Ord, 1979; Wilson et MacArthur, 1967). Cependant, la pratique de la conception systématique, explicite et délibérée de modèles qui a longtemps été préconisée (p. ex. Levins, 1966; Silvert, 1981) demeure aujourd'hui une pratique exemplaire. Fondamentalement, une bonne conception de modèle signifie comprendre et résoudre explicitement les compromis entre l'exactitude, la généralité et la précision du modèle.

Grâce à l'augmentation de la puissance de calcul, à la disponibilité des données et aux progrès continus des méthodes de modélisation, les MRE peuvent maintenant être établis de manière relativement simple. La mise en œuvre des modèles comporte toutefois de nombreuses étapes (c.-à-d. Figure 1.1.), et chaque étape (choix de la méthode, collecte des données, préparation des données, sélection des variables et sélection du modèle) est une source d'erreurs et d'incertitudes dans les prévisions finales des modèles. Les examens des modèles (GREGG et Chan, 2014; Planque *et al.*, 2011; Robinson *et al.*, 2017) révèlent que les aspects essentiels de la conception et de l'élaboration des modèles, ainsi que l'évaluation de l'incertitude, sont couramment négligés.

Étant donné la facilité avec laquelle les modèles peuvent maintenant être établis, ils sont généralement élaborés malgré l'inadéquation potentielle des données disponibles par rapport aux objectifs qu'ils visent. Des pratiques exemplaires sont donc justifiées et leur application est de plus en plus sollicitée (Araújo *et al.*, 2019; Grimm *et al.*, 2014). Araújo *et al.* (2019) proposent un ensemble complet de pratiques exemplaires concernant les données sur l'occurrence, les données prédictives, l'élaboration de modèles et l'évaluation de modèles, et proposent une échelle mobile allant de règles lacunaires à des règles d'or (largement ambitieuses). Bien qu'il ne suive pas strictement ces lignes directrices, ce travail porte une attention particulière à chaque catégorie, en évitant les pratiques lacunaires et en excellant dans les autres.

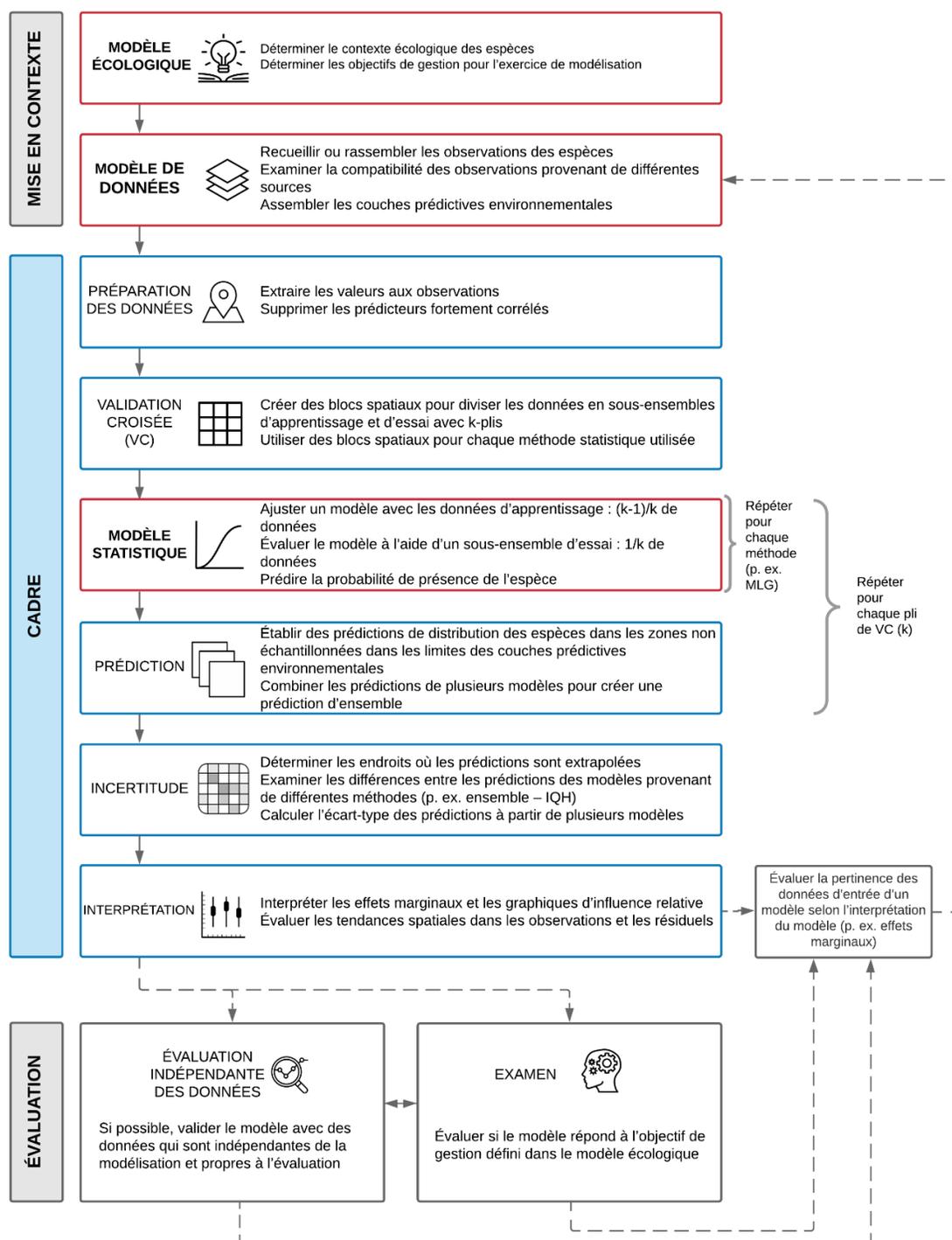


Figure 1.1. Aperçu du processus complet de modélisation, de l'élaboration du contexte du modèle (contextualisation) à l'évaluation des prédictions du modèle. Le cadre élaboré ici pour la modélisation de la répartition des espèces comprend une série d'étapes prescrites qui automatisent les pratiques exemplaires. Les modèles linéaires généralisés (MLG) sont un exemple de méthode de modélisation axée sur les données, tandis que les modèles d'indice de qualité de l'habitat (IQH) sont un exemple d'approche de modélisation d'enveloppe fondée sur les connaissances.

1.1.1 Le modèle écologique

La gestion d'une espèce ou de son habitat est indissociable de son écologie. Par conséquent, le modèle écologique décrit à la fois le contexte écologique et les aspects de gestion pertinents. La compréhension actuelle du cycle biologique, de l'environnement, des habitudes de déplacement et des niveaux d'exploitation de l'espèce fournit le contexte écologique dans lequel le modèle sera élaboré et les hypothèses qui devront l'étayer. Les aspects de gestion fournissent la justification de l'exercice de modélisation (p. ex. conservation, restauration, planification spatiale), ce qui permet d'éclairer les étendues et la résolution pertinentes. Un modèle écologique bien défini permet de définir un objectif de modélisation clair et d'éclairer le choix des étendues et de la résolution du modèle. Le contexte fourni par le modèle écologique est donc important pour définir le modèle de données.

Les caractéristiques des espèces influenceront également sur l'exactitude des prédictions qui en résulteront (p. ex. Elith *et al.*, 2002; Garrison et Lupo, 2002; Hepinstall *et al.*, 2002). Ces caractéristiques peuvent inclure l'état de conservation, le caractère distinctif de l'habitat, les mouvements locaux et le comportement migratoire, la taille de l'aire de répartition et le niveau trophique. Les études explorant les répercussions de ces caractéristiques ont donné des résultats différents (voir McPherson et Jetz, 2007), soulignant que le cycle biologique et les interactions biologiques potentielles peuvent influencer sur les résultats du modèle. Il est essentiel de comprendre les hypothèses nécessaires pour représenter les caractéristiques de ces espèces et les biais connexes dans les données d'observation pour en tirer des conclusions défendables et raisonnables.

1.1.2 Le modèle de données

Le modèle de données est composé des observations des espèces et des prédicteurs environnementaux (les variables qui caractérisent l'habitat de l'espèce). Les observations (c.-à-d. les données sur l'occurrence) s'accompagnent d'un contexte complet qui décrit comment, quand, pourquoi et où elles ont été recueillies. De même, le contexte des variables prédictives (les données indépendantes) indique leur résolution, si elles sont statiques (p. ex. l'altitude) ou supposées statiques (moyennes à long terme des observations), et toutes les interactions présumées entre elles. Différentes questions nécessiteront souvent que les variables prédictives dynamiques soient mises à différentes échelles.

Les données sur l'occurrence peuvent aller d'observations de la présence seulement, sans aucune information sur les lieux où l'espèce était absente, à des données sur l'abondance ou la densité obtenues par des relevés systématiques. Bien que les relevés systématiques soient considérés comme une pratique exemplaire (Araújo *et al.*, 2019), ils sont rarement mis en œuvre en raison de ressources limitées. Par conséquent, pour augmenter la taille de l'échantillon, les MRE sont souvent élaborés à partir d'assemblages de données provenant d'une diversité d'observateurs, de plateformes, de moments et de lieux, souvent recueillies au moyen de méthodes d'échantillonnage différentes. D'autres métadonnées importantes comprennent la précision temporelle et spatiale et la mesure dans laquelle les données sont liées à des stades biologiques particuliers (p. ex. larves ou adultes) ou associées à des événements saisonniers importants (p. ex. périodes optimales de croissance ou de frai). Cette information écologique est essentielle à l'interprétation des modèles, car les observations sont souvent regroupées en fonction de ces dimensions écologiques.

La disponibilité des variables prédictives et leur résolution (à la fois spatiale et temporelle) est le prochain facteur pratique à prendre en considération. Il s'agit d'un facteur déterminant clé des étendues spatiales, de la résolution et de la richesse écologique (au sens de Gregr et Chan, 2014) du modèle. Les prédicteurs environnementaux se situent sur un spectre de proximal à

distal, en fonction de leur position dans la chaîne des processus écologiques qui les relient à la répartition de l'espèce (Austin, 2002). L'utilisation de prédicteurs proximaux tend à produire des modèles plus robustes et généralisables, mais ces prédicteurs peuvent être difficiles à recueillir (Austin, 2002). Les variables distales reposent davantage sur des hypothèses concernant le processus, ce qui rend la validité de ces hypothèses primordiale pour l'efficacité du modèle pour les prévisions dans l'espace et le temps.

La sélection des variables est un aspect essentiel du modèle de données et, idéalement, elle est guidée par l'intersection des objectifs du modèle et de la disponibilité des données. Cependant, la disponibilité des données est souvent l'unique critère primordial. Il est donc important de s'assurer que les prédicteurs disponibles sont représentatifs à la fois de la zone d'intérêt et des processus écologiques présumés. Il est essentiel de comprendre les hypothèses sur la représentativité pour l'interprétation du modèle et l'évaluation de l'incertitude.

1.1.3 Le modèle statistique

Il existe de nombreuses façons de relier les espèces à leur environnement. Il existe des méthodes pour les espèces pour lesquelles il existe peu de données, voire aucune, des données sur la présence seulement, des données sur la présence et l'absence, et des données sur l'abondance ou la densité. Leur degré de complexité varie, allant de simples approches d'enveloppe ciblant les limites physiologiques des espèces à un éventail de méthodes fondées sur la régression et l'apprentissage machine. L'une des approches les plus simples (modèles IQH, p. ex. Kaschner *et al.*, 2006) repose entièrement sur les connaissances existantes. L'analyse factorielle des niches écologiques (Hirzel *et al.*, 2002) et BIOCLIM (examiné dans Elith *et al.*, 2006) sont d'autres méthodes de type enveloppe, dont la plupart utilisent des données seulement sur la présence. Lorsqu'on dispose d'observations de la présence ou de l'abondance (p. ex. les dénombrements) et de l'absence, on dispose d'une gamme d'approches de plus en plus exigeantes en données. Les méthodes de régression classiques (c.-à-d. les modèles linéaires généralisés, les modèles additifs généralisés et les modèles additifs généralisés mixtes) peuvent être utilisées pour prédire des formes fonctionnelles de plus en plus complexes, tandis que les méthodes d'apprentissage machine (p. ex. arbres de classification et de régression, réseaux neuronaux, entropie maximale) peuvent générer des relations plus complexes. Plus le modèle est complexe, plus il faut de données pour le paramétrer. Compte tenu de la diversité des méthodes et de leur complexité, le principal facteur qui entre dans le choix des méthodes de modélisation est la mesure dans laquelle elles répondent à l'objectif de gestion du modèle écologique. Dans la pratique, la disponibilité et la pertinence des données d'observation et des données prédictives (c.-à-d. le modèle de données) sont également des facteurs primordiaux à considérer.

Des études comparant la performance relative des différentes méthodes paraissent régulièrement (p. ex. Elith *et al.*, 2006; Guisan et Zimmermann, 2000; Pearson *et al.*, 2006; Segurado et Araújo, 2004). Bien que les études diffèrent en matière de détail, les méthodes de régression souples et certaines méthodes d'apprentissage machine tendent à être les meilleures pour expliquer les données d'observation. Toutefois, des approches plus simples (p. ex. modèles de niche ou d'enveloppe) peuvent suffire pour certaines applications, comme la prédiction de l'aire de répartition potentielle ou de ses déplacements possibles sur de grandes étendues spatiales (Cheung *et al.*, 2009), ou le soutien de contextes pauvres en données (Greegr *et al.*, 2018).

1.1.4 Intégration des connaissances écologiques locales

La description de la répartition des espèces peut bénéficier de l'intégration des connaissances écologiques locales (CEL) découlant de contacts réguliers et à long terme avec l'espèce et son

environnement, y compris les connaissances écologiques traditionnelles (CET) transmises de génération en génération. Ericksen et Woodley (2005) constatent que la pertinence, la crédibilité et la légitimité des modèles d'écosystème augmentent lorsque les connaissances, les besoins et les préoccupations de ceux qui sont principalement touchés par les résultats de la modélisation sont pris en compte. Des lignes directrices méthodologiques sur l'intégration des CEL aux modèles sont toujours en voie d'élaboration, mais un nombre croissant d'études sont menées sur la question (Belisle *et al.*, 2018). Pour intégrer les CEL, Belisle *et al.* (2018) recommandent une approche reproductible et multidisciplinaire qui incorpore des concepts des sciences sociales et de l'écologie pour traiter les données empiriques et les connaissances qui peuvent prendre la forme de mythes, de légendes ou de rites (p. ex. Colding et Folke, 2001).

Certes, le cadre et son application n'intègrent pas les CEL aux MRE, mais toute application future du cadre bénéficierait de leur inclusion, le cas échéant. Les CEL peuvent soutenir le processus d'élaboration du modèle à plusieurs étapes, depuis l'élaboration des modèles écologiques, statistiques et de données jusqu'à la validation indépendante du modèle. Les CEL peuvent être directement intégrées à un modèle d'enveloppe fondé sur les connaissances au cours du processus de sollicitation d'avis d'expert, et pourraient être combinées à des méthodes empiriques dans un cadre bayésien (Belisle *et al.*, 2018).

2 SÉLECTION DES DONNÉES

La présente section décrit les caractéristiques des données sur l'occurrence des espèces et des données prédictives qui peuvent entraîner des biais et l'introduction d'erreurs de prévision connexes, ce qui donne lieu à une piètre performance et à une mauvaise interprétation du modèle. Toutes les données s'accompagnent de biais et tous les modèles reposent sur des hypothèses quant à la pertinence des données aux fins du modèle. Par exemple, les observations sont souvent recueillies le long de voies de déplacement très fréquentées ou seulement dans une partie de l'aire de répartition ou de l'habitat réel de l'espèce. Le biais d'échantillonnage spatial qui en résulte peut faire en sorte que le modèle décrive les tendances aux sites échantillonnés plutôt que dans l'ensemble de l'aire de répartition de l'espèce (Barry et Elith, 2006). Les prédicteurs ne sont pas à l'abri d'erreurs ou de biais en raison du calcul de moyennes, des artefacts dans les données sources ou des méthodes d'interpolation qui peuvent entraîner des erreurs de prédictions locales (Barry et Elith, 2006). Les pratiques exemplaires consistent à prendre en compte ces biais dans les données sur l'occurrence lors de l'élaboration et de l'interprétation des modèles, une pratique qui s'applique ici.

Les données sur l'occurrence sont généralement obtenues sous forme de points ou de polygones, tandis que les données prédictives sont généralement représentées sous forme de grilles de points ou de matrices. Les relations entre les données sur l'occurrence et les données prédictives dépendront des étendues et de la résolution spatiale et temporelle des analyses (qui détermine comment la moyenne des observations et des prédicteurs est établie).

2.1 DONNÉES SUR L'OCCURRENCE

Les données sur les observations des espèces se présentent sous différentes formes et elles sont recueillies de différentes façons. Les observations les plus courantes, et les moins puissantes, sont des observations de la présence recueillies de façon non aléatoire, sans données crédibles sur l'absence. Néanmoins, l'abondance de ces ensembles de données a conduit à la mise au point de méthodes visant à faciliter l'utilisation de données sur la présence seulement dans l'élaboration des MRE (p. ex. Elith *et al.*, 2006; Phillips et Elith, 2010). Ce cadre de MRE met l'accent sur les observations recueillies au moyen de relevés conçus pour la surveillance, l'évaluation des stocks, ou la déclaration des prises. Toutes ces sources de

données contiennent (ou permettent d'estimer) des observations de l'absence, ajoutant ainsi des renseignements importants au processus de modélisation.

Il arrivera souvent que les observations enregistrent aussi l'abondance sous forme de dénombrements, d'estimations de la densité ou de captures par unité d'effort (CPUE). Bien que les données sur l'abondance soient essentielles pour calibrer les modèles de densité, elles sont également moins largement disponibles et exigent un traitement plus poussé des modèles. Le cadre se concentre donc sur les données binaires sur la présence-absence afin de permettre la prise en compte d'un plus large éventail de données et d'espèces. D'importants éléments du cadre sont néanmoins pertinents pour les données sur l'abondance.

En plus du type de données recueillies, d'autres caractéristiques des données sur l'occurrence peuvent influencer sur la capacité d'un modèle à décrire la répartition spatiale de l'espèce, et à discerner l'influence des prédicteurs.

2.1.1 Taille de l'échantillon

Le nombre d'observations utilisées pour la modélisation dépendra d'une variété de facteurs en raison du compromis entre la qualité et la quantité des données. La qualité des données dépend des méthodes de relevé, de la mesure dans laquelle elles ont été appliquées et de la durée des séries chronologiques. On peut élaborer des modèles avec moins de 100 observations (Hernandez *et al.*, 2006; Stockwell et Peterson, 2002), mais un échantillon de plusieurs centaines d'observations est plus habituel pour explorer efficacement la modélisation statistique. Des données de qualité sur l'absence et la prise en compte du nombre d'observations de présence et d'absence (c.-à-d. la prévalence) aboutiront à de meilleurs résultats du modèle (Barry et Elith, 2006; Lobo *et al.*, 2010).

Il conviendrait d'envisager la combinaison d'observations tirées de plusieurs relevés pour augmenter la taille de l'échantillon lorsque la taille de l'échantillon des observations de présence ou d'absence est faible ou lorsqu'un relevé ne couvre pas l'étendue de la modélisation. Toutefois, la combinaison de relevés et de différentes détectabilités d'espèces peut accroître l'incertitude du modèle en introduisant des biais d'échantillonnage multiples résultant de ces différences qu'il peut être difficile de dissocier.

2.1.2 Étendues

Les données d'observation les plus puissantes couvriront toute l'étendue spatiale de la répartition de l'espèce et incluront une gamme de différents habitats convenables. Cela revient à couvrir toute la gamme des conditions environnementales appropriées, ce qui est utile pour déterminer dans quelle mesure les résultats du modèle peuvent être appliqués dans l'espace et dans l'environnement. L'étendue temporelle des observations est également utile pour comprendre le contexte écologique, y compris si les données ont été recueillies sur une ou plusieurs années, ou pendant différentes saisons. Elle permet d'éclairer la généralisation des résultats à d'autres années et la mesure dans laquelle ils reflètent une période particulière de l'année (Gegr, 2011; Gegr *et al.*, 2018). Par conséquent, l'étendue temporelle des données sur l'occurrence influera sur les relations avec les prédicteurs dynamiques (p. ex. température, salinité) et peut mener à un ajustement du modèle à un ensemble anormal d'observations. De plus, si les observations ne couvrent pas toute l'étendue de l'aire de répartition de l'espèce, des descriptions importantes de l'habitat peuvent être omises (en raison de l'absence de données sur la présence réelle) ou mal différenciées (en raison du manque de données sur l'absence réelle).

2.1.3 Biais d'échantillonnage

Les meilleures pratiques exigent des observations découlant de relevés dont les plans d'échantillonnage sont statistiquement robustes et non biaisés, les observations étant distribuées dans toute l'aire de répartition de l'espèce. Des relevés non systématiques peuvent mener à des prédictions de tendances dans les méthodes de collecte des données plutôt que dans les associations d'habitats, surtout si l'échantillonnage est structuré sur le plan spatial ou environnemental (Araújo *et al.*, 2019). Par exemple, les données sur les prises dans les pêches proviennent habituellement de zones fortement propices à l'habitat, car les pêcheurs sont fortement incités à éviter les zones de faible densité d'espèces, ce qui mène souvent à l'hypothèse que les zones évitées représentent de véritables absences. Les observations peuvent aussi avoir une composante temporelle, notamment les changements saisonniers et annuels dans le comportement, la mobilité ou les stades du cycle biologique. Si les observations se limitent à un stade biologique particulier (p. ex. le frai) ou à une période de l'année (p. ex. la migration), ces limites doivent être explicitement prises en compte dans l'interprétation du modèle.

Enfin, le biais potentiel de disparition locale est un facteur de considération important pour les espèces exploitées. Si une espèce a été éliminée d'une partie donnée de son habitat, comme les eaux peu profondes, ou d'une région particulière en raison de la facilité de la récolte, et qu'aucune observation n'existe pour ces zones, les modèles corrélatifs pourraient ne pas permettre de saisir correctement les relations avec l'habitat. C'est particulièrement le cas si l'aire de répartition disparue présente des caractéristiques d'habitat uniques. Dans de tels cas, les modèles qui en résulteront sous-représenteront probablement l'habitat potentiel de l'espèce.

2.1.4 Précision

La précision avec laquelle les observations sont faites et enregistrées présente des dimensions spatiales, temporelles et taxonomiques. La précision spatiale des données sur l'occurrence devrait éclairer la résolution spatiale des variables prédictives, tandis que leur précision taxonomique présente un intérêt pour ce qui peut être déduit du MRE qui en résulte. De même, la précision temporelle est importante lorsque l'objectif de gestion se rapporte à des stades particuliers du cycle biologique ou à des différences saisonnières dans la qualité de l'habitat.

Ignorer la précision des données sur l'occurrence et utiliser aveuglément toutes les données disponibles est considéré comme une pratique lacunaire (Araújo *et al.*, 2019). C'est particulièrement vrai si les observations sont regroupées à partir de différentes méthodes de collecte, dont la précision peut varier dans l'espace, le temps ou la taxonomie. Les observations peuvent être faites à bord de navires ou sous l'eau, à distance ou directement, et ces facteurs ont des conséquences sur la précision avec laquelle elles peuvent être décrites.

2.2 DONNÉES PRÉDICTIVES

Les modèles d'habitat s'appuient sur des prédicteurs environnementaux pour décrire le monde abiotique. Les prédicteurs pertinents du milieu marin proviennent de diverses sources et peuvent être représentés sous forme de données continues ou catégorielles. Les technologies de télédétection (p. ex. imagerie satellitaire, drones et acoustique) et les modèles océaniques (y compris l'élévation, le type de fond et la dynamique des océans) constituent une source de données de plus en plus détaillées pour un nombre croissant de variables environnementales.

L'imagerie satellitaire produit des modèles de la température de la surface de la mer, des concentrations de chlorophylle-A, de la profondeur de la zone euphotique et de la vitesse et de la direction du vent. Les modèles de circulation océanique estiment la température, la salinité, la

profondeur de la couche de mélange et la vitesse et la direction du courant à diverses profondeurs. Lorsqu'ils sont combinés à un modèle biogéochimique, les modèles de circulation peuvent également prédire des propriétés biologiques et chimiques telles que les concentrations d'oxygène dissous, l'aragonite et le plancton. Ces prédicteurs dynamiques peuvent être représentés à différentes échelles temporelles (p. ex. saison, année, décennie). Le type de fond, un déterminant important de l'habitat pour de nombreuses espèces, peut être interpolé à partir des observations (GREGG *et al.*, 2013) ou prédit à l'aide de modèles (LI *et al.*, 2011). Enfin, la topographie représente une classe importante de prédicteurs couramment utilisés qui comprend la profondeur et une variété de dérivés tels que la pente, la courbure, l'indice de position bathymétrique (Walbridge *et al.*, 2018) et les mesures de la rugosité (p. ex. Du Preez, 2015; Sappington *et al.*, 2007).

Il est également possible d'inclure des variables spatiales, comme la distance par rapport à des caractéristiques physiques ou biotiques importantes (p. ex. le rivage, des récifs ou des peuplements de varech). Certains modèles utilisent la latitude et la longitude pour remplacer des prédicteurs inconnus, non représentés ou non représentables, mais l'interprétation de ces variables géographiques pose problème. De plus, l'effort de pêche peut être un prédicteur important de la répartition des espèces dans les écosystèmes touchés (p. ex. Foster *et al.*, 2015; Tien *et al.*, 2017).

2.2.1 Étendues et résolution

La fourchette des valeurs des variables prédictives utilisées varie généralement en fonction de la taille de la zone d'étude (p. ex. tenir compte de la gamme des profondeurs couvertes dans un modèle côtier par rapport à un modèle régional). De telles différences d'étendue peuvent influencer sur les relations mesurées avec les observations des espèces (Austin, 2007). Pour des étendues plus petites (c.-à-d. à l'intérieur de l'aire de répartition d'une espèce), les préférences d'une espèce peuvent ne pas être couvertes, tandis qu'une plus grande étendue peut permettre de déterminer les limites biologiques d'une espèce. Lorsque la fourchette d'un prédicteur dépasse de loin le caractère convenable de l'habitat de l'espèce et que les données sur l'absence ne sont pas disponibles, la pertinence biologique des relations peut être compromise (Austin, 2007; Fourcade *et al.*, 2018).

La résolution des prédicteurs marins disponibles peut couvrir plusieurs ordres de grandeur. Les données de télédétection et les modèles de circulation peuvent varier du kilomètre au mètre, tandis que les prédicteurs topographiques obtenus par des méthodes acoustiques peuvent être résolus à une échelle inférieure au mètre. Pour des raisons de calcul, la résolution est invariablement liée à l'étendue (les zones d'étude plus vastes nécessiteront une résolution réduite), et une résolution réduite entraîne une perte de précision puisque les variables sont moyennées en fonction de l'espace. Pour mieux prédire l'habitat d'une espèce, la résolution doit être pertinente à la façon dont l'espèce interagit avec son environnement (Wiens, 1989). Par exemple, la résolution spatiale des prédicteurs pertinents sera très différente pour un escargot de la zone intertidale et pour un cétacé qui migre sur de grandes distances. De même, comme les processus de l'habitat s'accomplissent à plusieurs résolutions (Levin, 1992), les prédicteurs agrégés à des résolutions quotidiennes, mensuelles ou saisonnières peuvent aider à saisir d'autres renseignements importants comparativement aux climatologies à long terme. Il est donc important d'examiner dans quelle mesure les prédicteurs correspondent à la résolution des observations, dans l'espace et dans le temps (Wiens, 1989). Par exemple, la rugosité dérivée d'une couche de profondeur à une résolution de 100 mètres peut ne pas être significative pour les espèces dont la relation avec la rugosité est connue à une échelle inférieure au mètre.

La résolution influe également sur la façon dont les données sont préparées et l'interprétation des résultats aux échelles locales. Par exemple, si une cellule de la grille n'est pas suffisamment échantillonnée, il peut en résulter des données d'absence erronées. De même, bien qu'une étude mondiale puisse prédire qu'une région côtière de 10 km x 10 km est très convenable, l'hétérogénéité du monde réel suggère qu'il est peu probable qu'une si grande superficie convienne de façon uniforme. Cela a des répercussions sur la mise à l'essai et l'interprétation des modèles.

2.2.2 Précision

Bien que les données des prédicteurs physiques recueillies en même temps que les observations soient les plus fiables, les prédicteurs modélisés permettent d'interpoler les prédictions des MRE sur l'ensemble de la zone d'intérêt. Les prédicteurs modélisés présentent inévitablement de façon inexacte certains aspects des caractéristiques réelles, ce qui entraîne des erreurs de représentation relatives (p. ex. fronts de marée où les apports d'eau douce ne sont pas inclus ou caractéristiques plus petites que la résolution du modèle) et absolues (p. ex. sédiments sablonneux prévus plutôt que vase). Si une telle fausse représentation est systématique, on peut la déceler en examinant la dérivée du prédicteur (p. ex. la courbure peut mettre en évidence des artefacts dans le prédicteur de la pente). Le calcul de la moyenne temporelle ou la mise à l'échelle peuvent causer des fausses représentations similaires (voir la section suivante). Cependant, tous les modèles présenteront des erreurs et, même avec des tests de validation puissants, la distribution spatiale de ces erreurs est difficile à établir. Bien que les pratiques exemplaires recommandent d'évaluer la façon dont les incertitudes liées aux prédicteurs influent sur les résultats du modèle (Araújo et al., 2019), on reconnaît qu'il s'agit là d'un défi important qui exige des méthodes de simulation sophistiquées. De telles simulations sortent du cadre de ces travaux. On reconnaît plutôt les incertitudes entourant l'exactitude des variables prédictives et on prend en compte leurs effets possibles sur l'interprétation du modèle.

2.2.3 Mise à l'échelle

Les variables prédictives existant à différentes résolutions spatiales, il faut mettre certaines à l'échelle de la résolution du modèle choisi. La pratique courante consiste à faire la moyenne de celles dont la résolution est plus fine que celle du modèle, tandis que les variables prédictives plus grossières peuvent être interpolées, rééchantillonnées ou réduites (c.-à-d. que la même valeur est appliquée à toutes les cellules de chaque unité d'étude plus grossière). Ces considérations relatives à l'échelle spatiale sont suffisantes pour les prédicteurs statiques (c.-à-d. ceux dont on suppose qu'ils ne varient pas avec le temps, comme la bathymétrie et ses dérivés).

La plupart des MRE dans la littérature sont encore présentés comme une seule carte statique des conditions moyennes. Cela signifie qu'il faut moyenniser tous les prédicteurs ayant une dimension temporelle (p. ex. les prédicteurs dynamiques comme la température, la salinité et la chlorophylle-A) dans le temps. La pratique exemplaire exige que ces valeurs moyennes reflètent la réponse biologique modélisée (Araújo et al., 2019), car la façon dont la moyenne d'un prédicteur est calculée peut influencer sur sa relation avec les observations de l'espèce (Eger et al., 2016; Levin, 1992; Wiens, 1989).

Ainsi, on fait souvent la moyenne des prédicteurs sur une série chronologique pour une période ou une durée pertinente pour le contexte écologique (les aspects écologiques ou de gestion du contexte écologique). Idéalement, la durée chevauchera également la période de la collecte des observations de l'espèce. La résolution temporelle et l'étendue du modèle peuvent être importantes pour la gestion, de façon saisonnière (p. ex. l'habitat de frai, ou encore l'occupation

estivale pour les espèces migratrices) ou annuelle (p. ex. la conception des aires protégées, la planification des interventions ou la gestion des pêches). La gestion dynamique des océans (Maxwell *et al.*, 2015) fournit un exemple de modèles de distribution à haute résolution temporelle. Il pourrait être utile d'inclure dans les modèles à plus haute résolution des événements stochastiques temporels, comme les zones mortes désoxygénées et les baisses de salinité, ou des événements mortels, comme les collisions avec des navires ou des contaminations.

Du fait de l'incohérence inévitable de la résolution entre les observations et les prédictors (et souvent entre les prédictors eux-mêmes), certaines corrélations seront probablement bien représentatives d'un processus écologique, mais d'autres serviront de valeurs de remplacement pour d'autres variables ou d'autres résolutions. Austin (2002) l'a caractérisée comme une distinction entre les variables proximales et distales, bien qu'en réalité, il s'agisse probablement davantage d'un spectre d'association entre observations et prédictors. L'importance de cette distinction dépendra des objectifs du modèle. Si les objectifs se rapportent à la compréhension du processus ou au transfert du modèle, des variables proximales sont souhaitables. Si l'objectif premier est d'expliquer la tendance observée dans la zone d'étude, les variables distales peuvent convenir également, même si elles peuvent être des approximations pour d'autres variables ou processus.

2.2.4 Prédictors spatiaux

L'autocorrélation spatiale (SAC) peut jouer un rôle important dans la répartition des espèces. Une SAC positive est souvent observée dans les données sur les espèces, que les observations soient binaires (présence/absence) ou continues (abondance). La dépendance spatiale entre les observations peut provenir de différentes sources, par exemple, des facteurs biotiques tels que la dispersion ou la présence de SAC dans les conditions physiques comme la température (Bahn et McGill, 2013). Il a été démontré que l'inclusion d'une mesure de la structure spatiale améliore la performance du modèle (Augustin *et al.*, 1996; Bahn et McGill, 2007; Martin *et al.*, 2014). Il existe plusieurs approches pour incorporer l'information sur la structure spatiale dans les MRE (voir Dormann *et al.*, 2007; Elith et Leathwick, 2009); cependant, les mérites des différentes approches demeurent un sujet de débat (Bini *et al.*, 2009; Hawkins, 2012; Kühn et Dormann, 2012). Leur complexité varie d'approches simples comme l'inclusion de coordonnées géographiques ou d'une autocovariable comme variables prédictives à des méthodes plus complexes utilisant des fonctions de covariance pour approximer des champs aléatoires spatiaux. Dans toute approche, il est important de tenir compte de l'échelle de la SAC par rapport à l'espacement de vos échantillons si vous utilisez des modèles spatiaux pour la prédiction, car les modèles spatiaux donneront de mauvaises prédictions si les écarts entre les échantillons sont plus grands que l'échelle de la SAC.

3 UN CADRE DE MODÉLISATION DE LA RÉPARTITION DES ESPÈCES

Cette section décrit le cadre de la modélisation de la répartition des espèces, élaboré d'après les pratiques exemplaires en matière de création de modèles et peaufiné grâce à l'expérience acquise au cours du processus d'application du cadre. Le cadre (Figure 1.1.) comprend six éléments : la préparation des données, la validation croisée, l'ajustement et l'évaluation des modèles, la prédiction, l'incertitude et l'interprétation. Il a été élaboré en conjonction avec des scripts spécialement conçus à cet effet et rédigés dans le langage de programmation statistique R (R Core Team, 2018). Le flux du cadre (Figure 3.1) peut être répété plusieurs fois avec différentes méthodes d'ajustement de modèle pour créer une prédiction de modèle d'ensemble (décrite dans la section 3.5). Les ensembles à plusieurs modèles produisent des prédictions

robustes (Oppel et al., 2011) et donnent un aperçu de l'incertitude du modèle (voir la section 3.6).

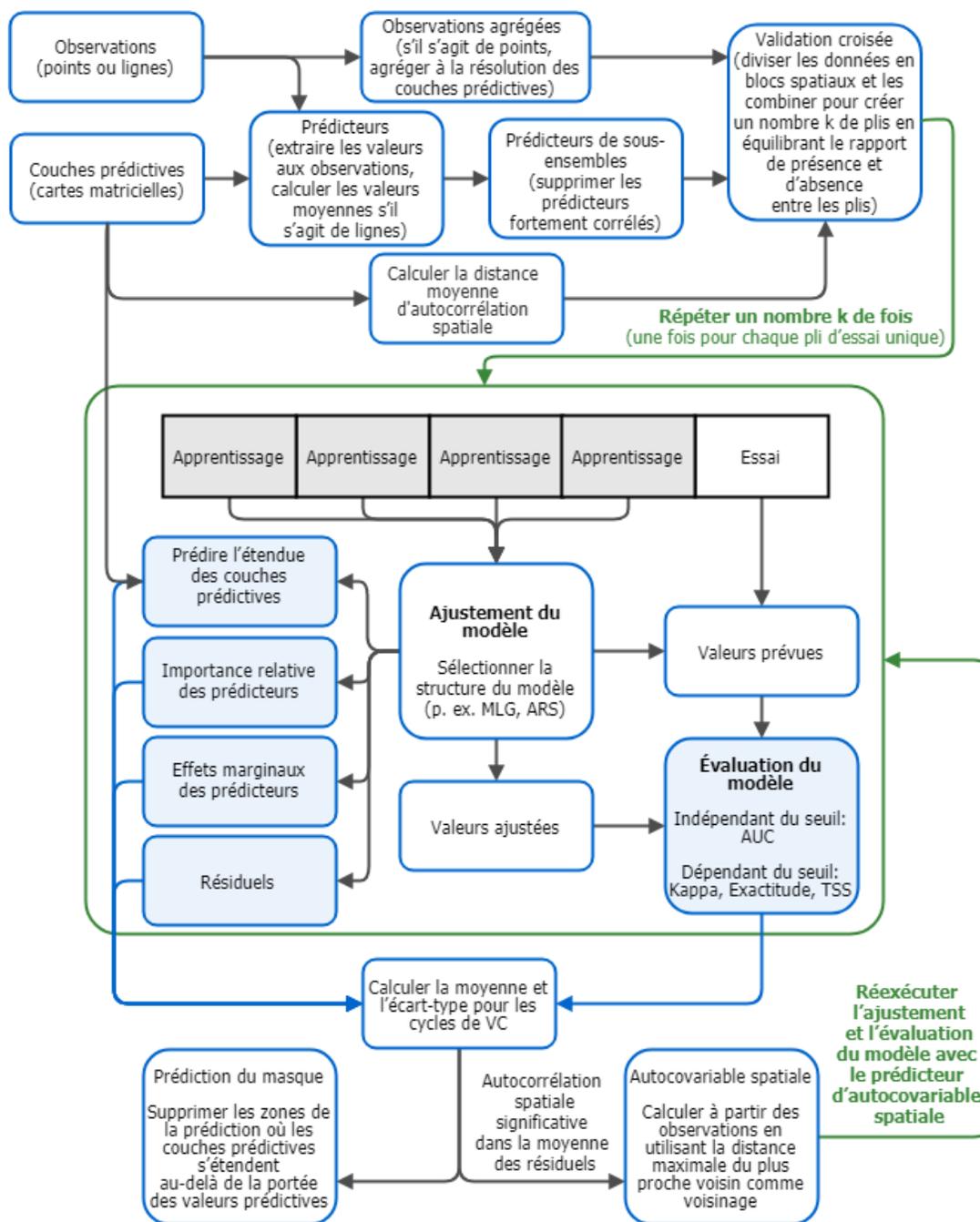


Figure 3.1. Diagramme du flux du cadre mis en œuvre avec des [scripts spécialement conçus à cet effet](#). Le flux comprend le traitement des observations (données dépendantes) et des couches de prédicteurs, ainsi que l'ajustement, l'évaluation et la prédiction des modèles selon une approche de validation croisée. L'approche d'ajustement du modèle varie selon le choix de la méthode de modélisation. Les modèles linéaires généralisés (MLG) et les modèles à arbre de régression stimulée (ARS) sont deux exemples de méthodes d'ajustement des modèles. Les cases bleues pleines représentent les extrants du processus de modélisation.

3.1 PRÉPARATION DES DONNÉES

Une fois que les données sur l'occurrence et les couches de prédicteurs ont été obtenues et qu'une résolution spatiale a été choisie, il y a plusieurs étapes à suivre pour préparer les données (compte tenu des considérations détaillées à la section 2) avant l'ajustement du modèle. Premièrement, il faut représenter les observations et les prédicteurs spatialement avec le même système de référence des coordonnées. Les données sur l'occurrence sont projetées de manière à correspondre à la projection géographique des prédicteurs. On vérifie ensuite les observations pour déceler les valeurs erronées (p. ex. les valeurs de présence-absence autres que 0 ou 1) ou manquantes, qui sont supprimées.

En général, les observations sont disponibles sous forme d'emplacements ponctuels (p. ex. des bennes) ou de segments de ligne (p. ex. la trajectoire du chalut). Les valeurs prédictives sont extraites des couches matricielles à l'emplacement des observations, mais le processus varie légèrement pour les données ponctuelles et linéaires. Pour les données ponctuelles, si plus d'une observation se situe dans une cellule matricielle prédictive, les points sont agrégés sur la grille matricielle. Cela permet de normaliser les observations et les prédicteurs à la même résolution spatiale. De plus, si l'on intègre les observations d'un même espace sur un certain nombre d'années, les observations sont mieux alignées dans le temps sur les couches prédictives qui représentent les conditions environnementales moyennes à long terme. Pour les observations de présence-absence, l'agrégation est effectuée en attribuant une présence à la cellule s'il y a au moins une observation de présence dans la cellule (d'après Guinotte et Davies, 2014). Avec les données de présence/absence, cette étape peut conduire à des données biaisées en faveur de la présence. Il est donc important de se demander si la modélisation de la probabilité d'occurrence à l'aide de cette méthode d'agrégation correspond aux objectifs des modèles (p. ex. la modélisation de la probabilité de présence à un moment dans le passé). Une fois les observations agrégées, on calcule le changement de taille de l'échantillon qui en résulte. Si le nombre d'observations a changé de façon marquée (p. ex. réduction de 50 %), il y a un décalage évident entre la résolution des observations et les variables environnementales, ce qui indique que la résolution choisie ne convient peut-être pas à la modélisation.

Pour les observations représentées sous forme de lignes, les valeurs moyennes sont calculées pour chaque cellule matricielle qui croise les segments de ligne (d'après Carrasquilla-Henao *et al.*, 2018). Pour les variables prédictives catégorielles, on utilise la catégorie la plus dominante qui se trouve dans les cellules matricielles intersectées. Cette approche intègre les valeurs prédictives pour l'ensemble du segment de ligne au lieu de simplement extraire une valeur unique au début, au milieu ou à la fin du segment de ligne. Après avoir préparé les observations pour la modélisation, on les cartographie et il est possible de les examiner visuellement pour déceler les erreurs de géoréférencement (Figure 3.2).

Une fois que les valeurs prédictives ont été extraites des couches matricielles vers les lieux d'observation, plusieurs contrôles de la qualité des données sont nécessaires. Les valeurs manquantes dans l'ensemble de données prédictives – ce qui peut se produire lorsque des observations ne s'inscrivent pas dans l'étendue des couches matricielles – sont supprimées. Le nombre final d'enregistrements dans les ensembles de données représente maintenant la taille de l'échantillon de l'ensemble complet des observations et des prédicteurs utilisés pour la modélisation. On examine ensuite si les prédicteurs conviennent au modèle en comparant la distribution de toutes les valeurs dans les couches matricielles aux valeurs des variables prédictives extraites. Idéalement, la plage des valeurs extraites devrait correspondre à celle de la matrice source qui sera utilisée ultérieurement pour la prédiction. Si la plage est tronquée, les prédictions du modèle dans certaines zones seront extrapolées. Des cartes sont produites pour chaque couche de prédicteurs afin de permettre une vérification visuelle des artefacts

apparents sur les valeurs inattendues. Le facteur d'inflation de la variance (VIF) et la corrélation de Spearman sont utilisés pour évaluer la colinéarité entre les variables prédictives extraites. Des valeurs de VIF supérieures à 10 peuvent indiquer qu'un prédicteur est fortement corrélé avec au moins un autre prédicteur et qu'il faudrait envisager de le supprimer (Dormann *et al.*, 2013). Toutefois, il n'est pas toujours nécessaire ou souhaitable de supprimer les prédicteurs corrélés. Même si un prédicteur fortement corrélé est supprimé avant la modélisation, il est difficile de déterminer l'influence de ses prédicteurs corrélés dans le modèle, car on ne peut pas démêler leurs effets relatifs sur la variable dépendante.

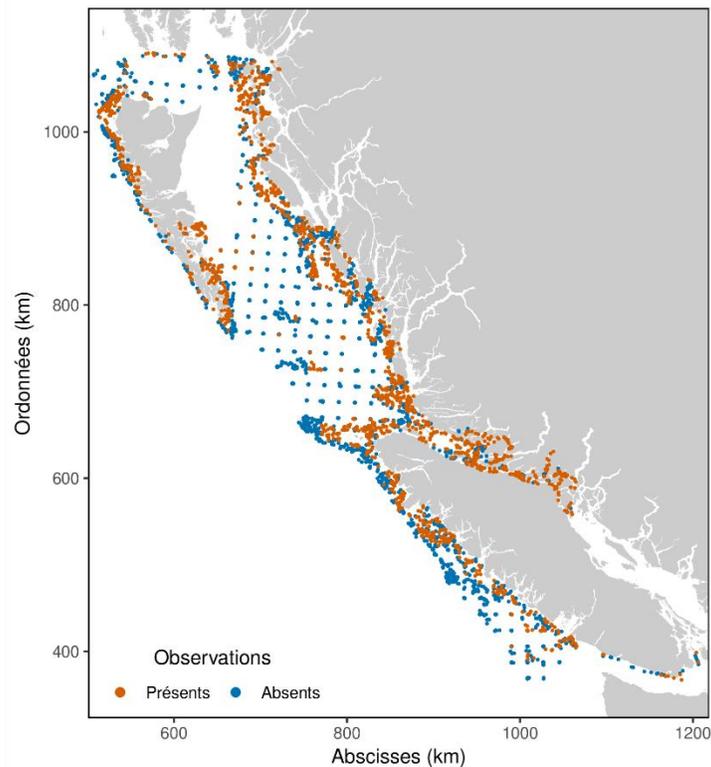


Figure 3.2. Distribution des observations de présence et d'absence d'une espèce dans une zone d'étude.

3.2 VALIDATION CROISÉE

Une fois préparé, le modèle de données est divisé en ensembles de données d'apprentissage et d'essai, utilisés respectivement pour l'ajustement et la validation du modèle. Il est essentiel de séparer les données d'essai et d'apprentissage pour évaluer la performance prédictive du modèle (Hijmans, 2012). La validation croisée (VC) est une procédure (Roberts *et al.*, 2016) largement utilisée qui sépare les données d'apprentissage et d'essai de manière à permettre l'utilisation de toutes les données dans le processus d'ajustement du modèle (Wenger et Olden, 2012). La VC fonctionne en divisant les données disponibles en k sous-ensembles (ou « plis » selon la méthode k -plis). Pour chaque exécution (cycle) de la VC, un pli différent est réservé pour les tests, et les autres sont utilisés pour l'entraînement. Le processus est répété k fois jusqu'à ce que chaque pli ait été utilisé une fois pour les tests. Ainsi, pour une VC à cinq plis, on élabore cinq modèles, chacun avec un $4/5^{\text{e}}$ différent des données et testé par rapport au $1/5^{\text{e}}$ restant. Les plis peuvent être divisés aléatoirement ou en créant des blocs. Les blocs peuvent être spatiaux ou créés en fonction d'autres facteurs dont on sait qu'ils peuvent limiter l'indépendance des données (voir Roberts *et al.*, 2016 pour des exemples). La VC aléatoire est encore utilisée pour évaluer les MRE bien qu'on sache qu'elle gonfle les scores de performance

parce que les séparations n'augmentent pas l'indépendance (Roberts *et al.*, 2016). L'indépendance des données d'apprentissage et d'essai est une hypothèse fondamentale qui sous-tend les tests statistiques de performance des modèles (Legendre, 1993). Les blocs spatiaux (recommandés et mis en œuvre dans ce cadre) améliorent l'indépendance spatiale des données d'apprentissage et d'essai, ce qui permet d'estimer plus précisément la performance et la transférabilité du modèle (Fourcade *et al.*, 2018; Hijmans, 2012; Merow *et al.*, 2014; Roberts *et al.*, 2016; Trachsel et Telford, 2016).

L'utilisation de la VC par blocs spatiaux pour accroître l'indépendance entre les plis est considérée comme une pratique exemplaire pour les observations de sous-échantillonnage aux fins de l'ajustement et de l'évaluation des modèles. Ce cadre applique la procédure de blocs spatiaux mise au point par Valavi *et al.* (2018), dans laquelle la taille optimale des blocs est déterminée par l'étendue de l'autocorrélation spatiale dans les couches prédictives. Les blocs sont assignés aléatoirement aux plis, avec plusieurs blocs dans chaque pli. Le processus est répété pour 250 itérations et on utilise les plis qui possèdent la taille d'échantillon la plus uniformément dispersée d'observations de présence et d'absence (Figure 3.3).

L'ajustement et la validation des modèles sont effectués plusieurs fois, une fois pour chaque cycle VC. On n'obtient pas un seul meilleur modèle, mais on choisit plutôt plusieurs meilleurs modèles, chacun adapté aux données d'apprentissage de son exécution de VC respective. Ainsi, les résultats (c.-à-d. les prédictions, les paramètres de validation, l'influence relative des prédicteurs et les effets marginaux) sont moyennés et on calcule l'écart-type ou la fourchette pour illustrer la variation entre les modèles de VC.

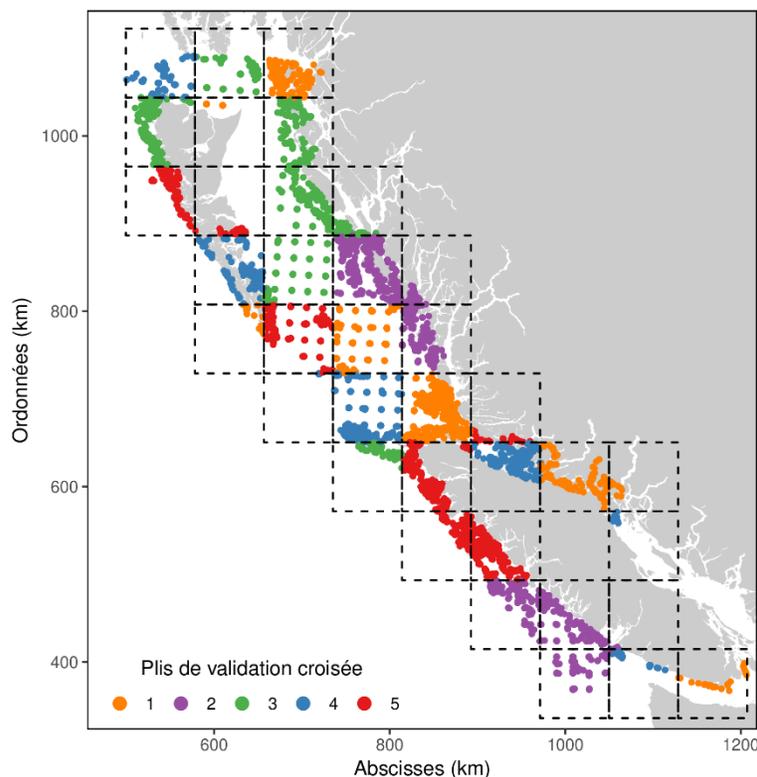


Figure 3.3. Orientation spatiale de la validation croisée par blocs spatiaux à cinq plis d'une espèce dans une zone d'étude. Chaque pli serait utilisé une fois comme données d'essai pour évaluer les modèles élaborés avec les données d'apprentissage constituées des plis restants.

3.3 AJUSTEMENT DU MODÈLE

L'ajustement d'un modèle pour établir une relation entre les prédicteurs environnementaux et les observations des espèces est l'exercice technique central de l'élaboration d'un MRE. L'ajustement des modèles peut aller de simples approches de modèles d'enveloppe fondés sur les connaissances (p. ex. IQH) qui n'exigent pas de données sur la présence des espèces et reposent sur des relations publiées et des consultations d'experts, à des algorithmes statistiques complexes (p. ex. ARS) qui permettent de discerner les corrélations entre les prédicteurs et les données sur les espèces. Ce cadre prend en compte l'ajustement pour inclure à la fois la structure du modèle (sélection des prédicteurs appropriés et de leur forme) et son paramétrage (détermination des paramètres qui permettent à la structure du modèle d'expliquer au mieux la tendance dans les observations). Pour les approches fondées sur les connaissances, la structure du modèle peut prendre les formes suivantes : linéaire, seuil, optimal, exponentiel ou logarithmique (Figure 3.4). Le paramétrage peut être progressif et itératif. Lorsque l'on utilise des méthodes corrélatives axées sur les données (p. ex. MLG ou ARS), l'établissement de la structure et le paramétrage ont lieu simultanément.

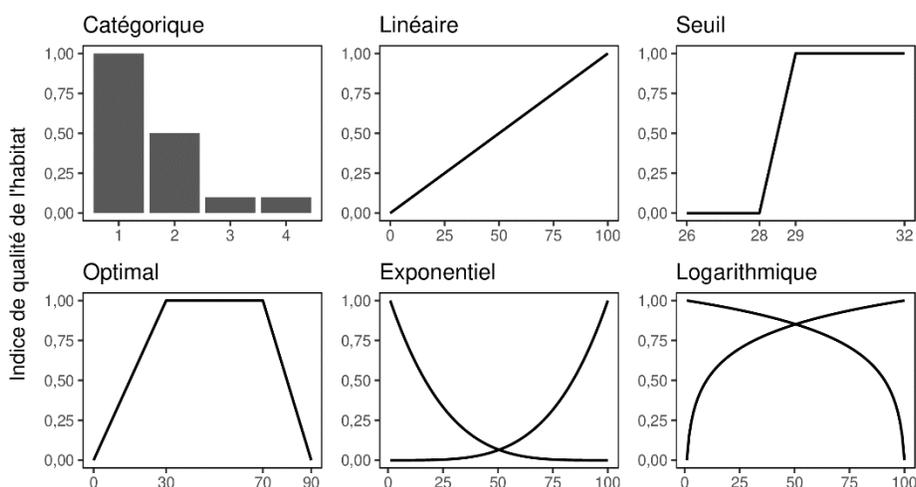


Figure 3.4. Exemples de relations possibles entre les variables prédictives environnementales et la qualité de l'habitat utilisées pour élaborer des modèles d'indice de qualité de l'habitat.

Le but de l'ajustement du modèle est d'expliquer autant que possible la variance dans les observations, sans surajuster le modèle. Un modèle est surajusté lorsque sa complexité (p. ex. le nombre de paramètres) est supérieure à ce qui est nécessaire pour représenter avec précision la relation entre les observations et les prédicteurs. Un tel modèle correspondra trop étroitement aux observations d'apprentissage et donnera de piètres résultats lorsqu'il sera appliqué aux observations d'essai. Il est entendu qu'il en résulte une perte de généralité, ce qui peut conduire à une mauvaise transférabilité du modèle (Wenger et Olden, 2012). Les modèles d'enveloppe fondés sur les connaissances ne peuvent techniquement pas être surajustés, car ils n'utilisent pas de données. Ils peuvent toutefois être surspécifiés pour une zone donnée où l'opinion d'experts locaux est prise en compte.

Les approches de modélisation utilisent diverses statistiques pour équilibrer le sous-ajustement et le surajustement (Elith et Leathwick 2009; Merow *et al.*, 2014). Pour les modèles de type régressif (p. ex. MLG), le critère d'information d'Akaike (AIC) est couramment utilisé comme méthode quantitative et non subjective pour choisir le meilleur modèle parmi un certain nombre de modèles candidats en pénalisant la complexité de l'ajustement du modèle (Burnham et Anderson, 2004). Dans le présent cadre, nous utilisons le critère d'information bayésien (BIC)

pour sélectionner le meilleur modèle parce qu'il applique une plus grande pénalité pour la complexité que l'AIC (Galipaud *et al.*, 2014; Link et Barker, 2006; Schwarz, 1978), ce qui donne des modèles plus simples. Une autre méthode d'ajustement de modèle (non utilisée dans le présent cadre) est la régularisation. Comme le BIC, la régularisation pénalise la complexité, mais elle le fait en réduisant les valeurs des coefficients, ce qui fait que les coefficients des prédicteurs moins influents se rapprochent de zéro (Elith *et al.*, 2011). Pour les méthodes d'apprentissage machine (p. ex. ARS), il est possible d'extraire une partie des données de l'ajustement du modèle aux fins du calcul d'une condition d'arrêt (Hastie *et al.*, 2009). Pour les ARS, on le fait en évaluant de manière itérative la performance du modèle à l'aide des données d'exclusion à intervalles réguliers pendant l'ajustement du modèle. Au fur et à mesure que la complexité du modèle augmente, les performances mesurées par rapport aux données d'exclusion commenceront à diminuer, et une condition d'arrêt sera déclenchée.

Il est crucial que les données d'essai restent indépendantes de l'ajustement du modèle si l'on veut obtenir des mesures précises de la performance (Araújo *et al.*, 2019; Hawkins *et al.*, 2003; Roberts *et al.*, 2016). Par conséquent, le cadre ne recommande pas d'étudier les relations univariées entre les observations et chaque prédicteur avant de séparer les données d'essai et d'apprentissage. Cette étape est souvent entreprise pour aider à décider des prédicteurs à inclure dans le modèle; cependant, elle contredit l'hypothèse d'indépendance entre les données d'essai et l'ajustement du modèle et peut mener à des mesures gonflées de la performance du modèle (Hastie *et al.*, 2009; Hawkins *et al.*, 2003) et à des modèles surajustés (Burnham et Anderson, 2004). Le cadre favorise plutôt la sélection des variables en deux étapes.

Premièrement, les prédicteurs sans relation écologique connue ou théorisée avec l'espèce modélisée sont éliminés par filtrage avant la modélisation. Deuxièmement, pour les modèles corrélatifs, les prédicteurs qui décrivent mal la distribution des observations sont éliminés par filtrage au cours de la procédure d'ajustement du modèle. Pour éliminer les prédicteurs par filtrage, on peut soit les retirer complètement de la structure du modèle (p. ex. lors de la sélection du modèle MLG avec le critère BIC), soit en les laissant dans la structure du modèle, mais de manière à ce qu'ils contribuent peu à la variance globale expliquée (p. ex. ARS).

3.4 ÉVALUATION DU MODÈLE

Il est possible d'évaluer l'ajustement du modèle et la performance prédictive pour les MRE. On évalue l'ajustement du modèle en mesurant la variance restant dans les données utilisées pour élaborer le modèle, tandis que la performance prédictive est mesurée d'après l'accord entre les prédictions du modèle et les observations non utilisées dans l'ajustement du modèle. La comparaison des statistiques mesurant l'ajustement du modèle et la performance prédictive peut fournir une estimation qualitative du surajustement. Des valeurs de performance plus élevées pour les données d'apprentissage par rapport aux données d'essai dénotent un modèle surajusté (Rooper *et al.*, 2017; Wenger et Olden, 2012). L'évaluation de la performance d'un modèle à l'aide de données indépendantes de l'ajustement du modèle est une étape nécessaire et importante de tout exercice de MRE, qu'il s'agisse de prédiction ou d'explication (p. ex. la formulation d'hypothèses concernant les facteurs environnementaux de la répartition des espèces).

Il existe une variété de paramètres pour mesurer la performance, y compris des statistiques indépendantes et dépendantes du seuil. Les statistiques indépendantes du seuil calculent la relation entre les observations et les prévisions continues (p. ex. probabilité d'occurrence) et sont recommandées comme principal paramètre de la performance du modèle pour évaluer les prédictions continues du modèle (Lawson *et al.*, 2014). Les statistiques dépendantes du seuil exigent que les prédictions continues soient converties en prédictions de présence-absence à l'aide d'une valeur seuil déterminée. Ces statistiques sont recommandées lorsque les prédictions de présence-absence sont l'objectif final.

Le présent cadre adopte un paramètre couramment utilisé, l'aire sous la courbe de fonction d'efficacité du récepteur (AUC), comme mesure de la performance indépendante du seuil (Elith *et al.*, 2006). La valeur AUC varie de 0 à 1. Les valeurs inférieures à 0,5 indiquent que les modèles sont pires que les modèles aléatoires, les valeurs de 0,5 indiquent que le modèle n'est pas meilleur que les modèles aléatoires, et les valeurs de 1 indiquent que le modèle prédit parfaitement les données d'occurrence (Freeman et Moisen 2008; Merckx *et al.*, 2011). Selon Pearce et Ferrier (2000) et Jones *et al.* (2010), les valeurs AUC supérieures à 0,9 sont considérées comme bonnes, comme moyennes entre 0,7 et 0,9, et comme faibles en dessous de 0,7.

À titre de comparaison, comme l'utilité de la statistique AUC a été remise en question dans des ouvrages (Lobo *et al.*, 2008), plusieurs paramètres dépendants du seuil sont également calculés : la sensibilité, la spécificité, la *true skill statistic* (TSS), le kappa et l'exactitude. La TSS équilibre la sensibilité (proportion des observations de présence qui sont correctement classées) et la spécificité (proportion des observations d'absence qui sont correctement classées) et est indépendante de la prévalence des observations (Allouche *et al.*, 2006). Le kappa est une mesure de l'accord entre les valeurs observées et prédites, qui tient compte des accords fortuits et dépend de la prévalence des observations. La TSS et le kappa vont de -1 à 1, les valeurs inférieures à 0 représentant des modèles qui ne sont pas meilleurs que les modèles aléatoires et les valeurs de 1 indiquant un accord parfait (Allouche *et al.*, 2006). Les valeurs de la TSS supérieures à 0,6 sont considérées comme bonnes, comme modérées entre 0,2 et 0,6 et comme faibles en dessous de 0,2 (Jones *et al.*, 2010; Landis et Koch, 1977). L'exactitude est le pourcentage des prédictions qui sont correctement classées et varie de 0 à 1, où 1 est l'exactitude la plus élevée.

Pour calculer les paramètres dépendants du seuil, le cadre utilise les pratiques exemplaires recommandées et détermine un seuil optimal en évaluant la performance du modèle à l'aide d'une statistique telle que la TSS sur plusieurs seuils, au lieu d'appliquer une valeur par défaut de 0,5 (Freeman et Moisen, 2008). La statistique idéale (p. ex. TSS, kappa ou exactitude) pour l'optimisation du seuil dépendra de l'objectif de la modélisation. Par exemple, une statistique qui favorise la spécificité plutôt que la sensibilité peut être appropriée pour certaines applications de gestion qui nécessitent un faible taux de faux positifs (Freeman et Moisen, 2008).

Plusieurs courbes de validation sont produites à l'aide de statistiques dépendantes et indépendantes du seuil, comme les courbes de fonction d'efficacité du récepteur (ROC) sur lesquelles l'AUC est basée (Figure 3.5), les courbes qui montrent des statistiques dépendantes du seuil calculées sur une gamme de valeurs seuils (Figure 3.6), et une matrice de contingence graphique illustrant la distribution de densité des prédictions de positifs vrais et faux et des prédictions de négatifs vrais et faux (Figure 3.7). L'examen de ces courbes, bien que qualitatif, est un aspect important de l'évaluation du modèle. Les courbes de validation peuvent être plus informatives qu'une seule statistique parce que les statistiques peuvent être examinées pour une gamme de seuils et que la densité des observations qui se situent au-dessus et au-dessous des seuils de présence-absence peut être visualisée. Les courbes de validation peuvent également servir pour effectuer des contrôles de qualité en vérifiant que les courbes correspondent aux statistiques connexes (p. ex. la courbe ROC correspond à la statistique AUC).

Enfin, les courbes de validation des résiduels sont produites sous forme de diagramme de dispersion des résiduels par rapport à chaque variable prédictive. Ces courbes peuvent aider à déterminer dans quelle mesure la structure du modèle est appropriée pour les données : si on observe une forte tendance résiduelle, il reste une variation inexplicée qui pourrait être mieux résolue avec une structure de modèle différente.

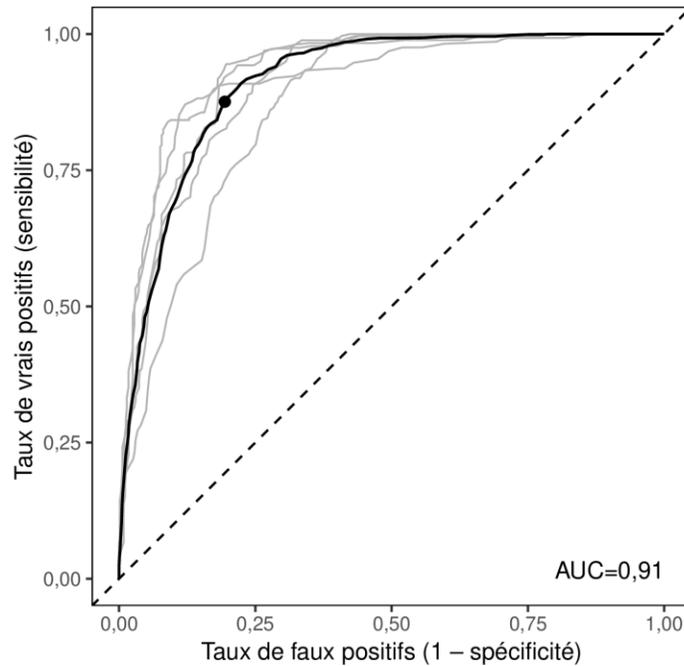


Figure 3.5. Courbe de fonction d'efficacité du récepteur (ROC). Les lignes grises représentent les courbes ROC des modèles à validation croisée à cinq plis et la ligne noire représente la moyenne des courbes ROC. Les courbes sont établies à partir des données d'essai. AUC désigne l'aire sous la courbe ROC. Le point noir indique le taux de faux positifs et de vrais positifs au seuil optimisé.

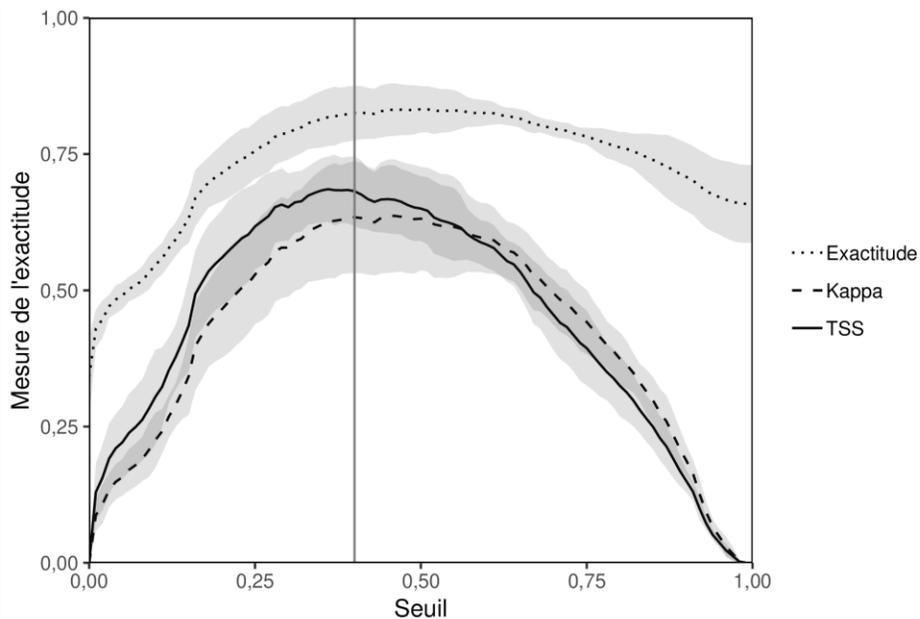


Figure 3.6. Courbes représentant les mesures de l'exactitude d'un modèle dépendant du seuil sur une gamme de seuils. Les courbes sont construites à partir des données d'essai. Les rubans gris représentent un écart-type autour de la moyenne calculée à partir de modèles à VC à cinq plis. TSS désigne la « true skill statistic ».

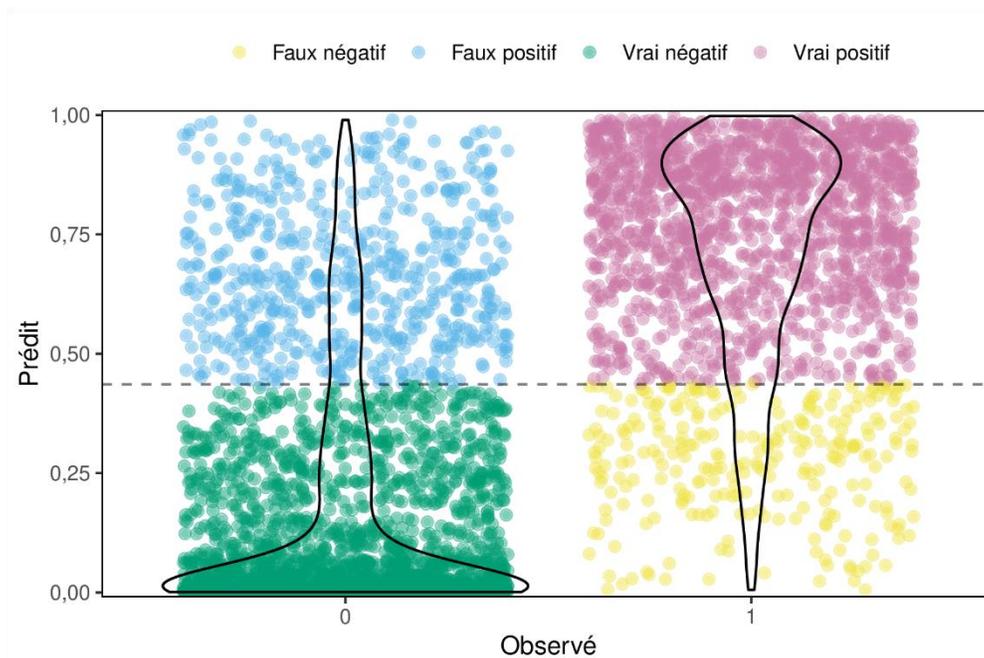


Figure 3.7. Représentation visuelle d'une matrice de contingence montrant la densité des prédictions de positifs vrais et faux au-dessus du seuil et des prédictions de négatifs vrais et faux au-dessous du seuil. Le seuil est représenté par une ligne pointillée grise. Les polygones noirs sont des estimations de la densité du noyau qui illustrent la densité des prédictions correspondant aux observations de présence et d'absence.

3.5 PRÉDICTION

Une fois qu'un modèle a été évalué, l'objectif des MRE est généralement de produire des prédictions pour les zones non échantillonnées de la zone d'étude où les valeurs prédictives sont connues. Dans ce cadre, des prédictions sont produites à partir de modèles tirés de chaque cycle VC (selon le flux détaillé à la Figure 3.1). Ce flux peut ensuite être répété avec différentes méthodes d'ajustement de modèle, ce qui permet d'élaborer une prédiction de modèle d'ensemble pour chaque cycle VC (Figure 3.8). Les prédictions sont combinées par le biais d'une moyenne pondérée en utilisant l'AUC pour la pondération (Anderson *et al.*, 2016; Opper *et al.*, 2011). Avant la pondération, les valeurs AUC comprises entre 0,5 et 1 sont remises à l'échelle de 0 à 1, les valeurs AUC inférieures à 0,5 étant fixées à 0. Les pondérations modifiées de l'AUC font en sorte qu'un modèle dont les résultats sont inférieurs ou égaux à ceux du modèle aléatoire ($AUC \leq 0,5$) ne contribue pas à la prédiction d'ensemble. Ce critère d'entrée du modèle dans l'ensemble a été choisi, car toute valeur plus élevée (p. ex. 0,7) serait arbitraire. Par la suite, la moyenne des prédictions d'ensemble est calculée et utilisée comme prédiction finale de l'ensemble (Figure 3.8).

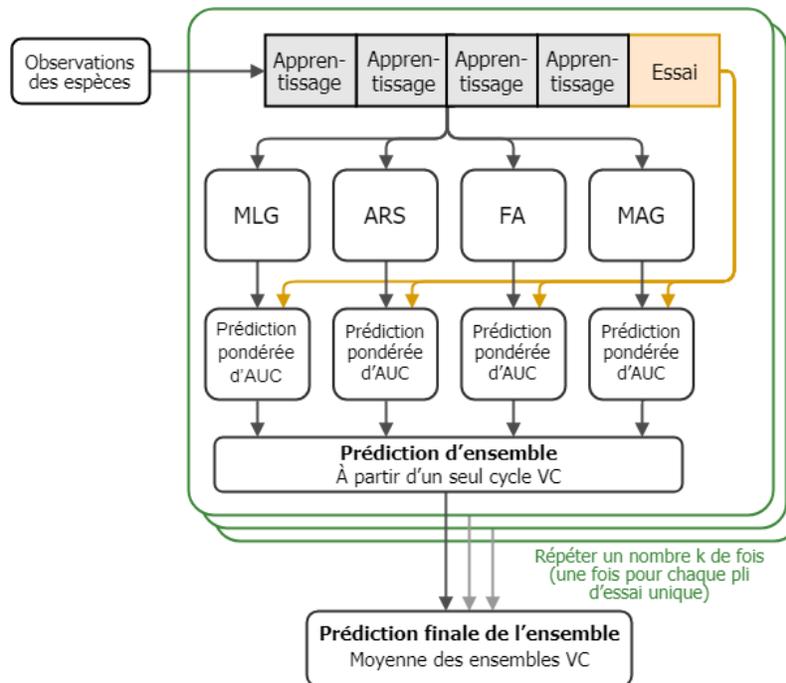


Figure 3.8. Diagramme du flux du modèle d'ensemble montrant comment les prédictions issues de différentes méthodes de modélisation peuvent être combinées en un ensemble dans la structure de validation croisée. Chaque modèle, y compris l'ensemble, est validé avec un jeu identique de données d'essai pour chaque procédure de validation croisée. Les modèles linéaires généralisés (MLG), l'arbre de régression simulée (ARS), la forêt d'arbres décisionnels (FA) et les modèles additifs généralisés (MAG) sont quatre exemples de méthodes d'ajustement des modèles.

En plus de la prédiction d'ensemble, le cadre recommande de produire un modèle fondé sur les connaissances à partir des connaissances spécialisées disponibles sur les relations entre une espèce et les conditions environnementales. De tels modèles d'enveloppe servent à illustrer la façon dont les connaissances spécialisées sur le processus peuvent être représentées et fournissent un point de référence auquel comparer la prédiction du modèle d'ensemble fondé sur les données (Figure 3.9).

La question de savoir s'il convient d'intégrer un modèle d'enveloppe dans l'ensemble, plutôt que de l'utiliser comme modèle de référence aux fins de comparaisons, dépend des considérations examinées dans la section 1.1, par exemple si les objectifs du modèle sont adaptés. Une pratique exemplaire importante à cet égard consiste à s'assurer que, lorsque les prédictions d'un modèle sont combinées en un ensemble, elles représentent les mêmes unités. Par exemple, il ne serait pas approprié de combiner des prédictions de l'abondance avec des prédictions de la probabilité d'occurrence. Le cadre n'inclut pas les modèles d'enveloppe dans l'ensemble, mais appuie leur utilisation pour explorer les aspects de l'incertitude du modèle (voir la section 3.6).

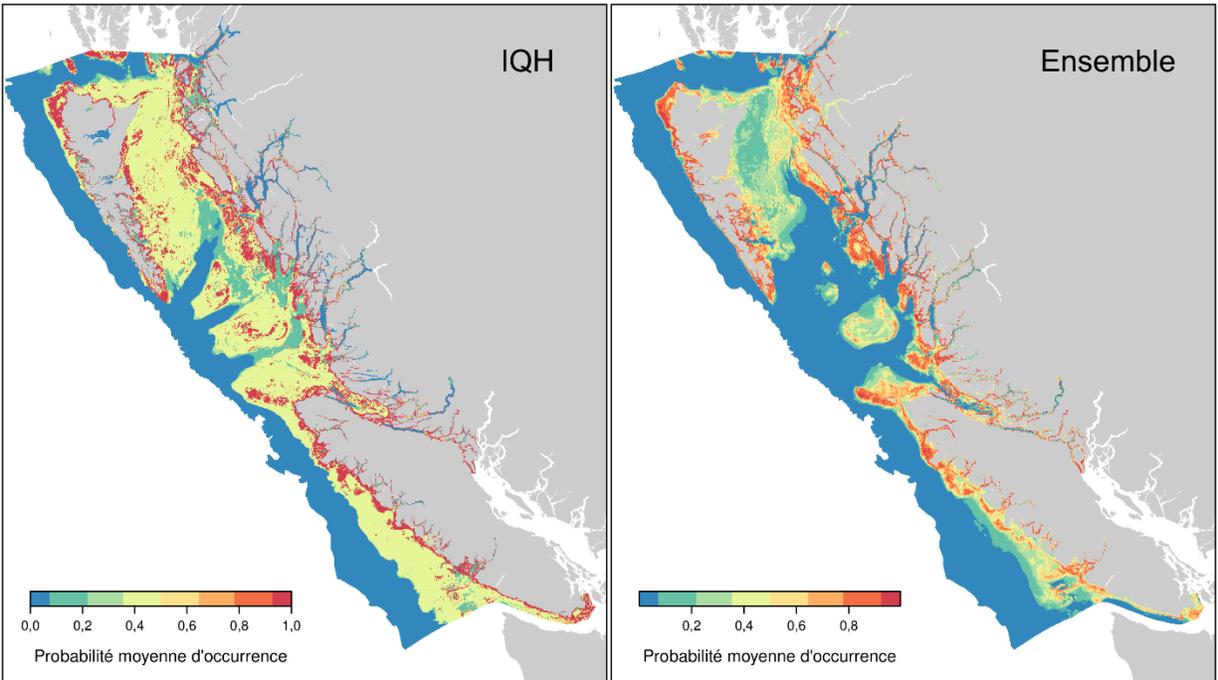


Figure 3.9. Prédications de la probabilité d'occurrence à partir d'un modèle d'enveloppe fondé sur les connaissances (p. ex. IQH, à gauche) et d'un modèle d'ensemble axé sur les données (à droite). La prédiction d'ensemble est la moyenne des prédictions des modèles à validation croisée à cinq plis élaborés à partir de méthodes multiples (p. ex. MLG et ARS).

3.6 INCERTITUDE

Une fois que les modèles ont été utilisés pour produire des prédictions sur les secteurs non échantillonnés de la zone d'étude, l'étape suivante consiste à examiner le degré d'incertitude de ces prédictions. L'incertitude des modèles peut provenir de nombreuses sources telles que la qualité et la disponibilité des connaissances spécialisées, les données sur l'occurrence et les données environnementales, les biais d'échantillonnage spatial et temporel, l'estimation des paramètres et de la structure des modèles, les méthodes statistiques utilisées et la variabilité naturelle du système (Barry et Elith, 2006; Dormann *et al.*, 2008; Iturbide *et al.*, 2018; Link *et al.*, 2012). Il est essentiel d'estimer l'incertitude des prédictions pour guider efficacement les décisions stratégiques telles que les fermetures des pêches ou la planification spatiale en mer (Jones-Farrand *et al.*, 2011; Jones et Cheung, 2015), quoique l'incertitude des prédictions des modèles soit rarement évaluée dans les applications marines des MRE (Robinson *et al.*, 2017). Ce cadre présente quatre méthodes permettant de réduire ou de décrire l'incertitude dans l'espace dans la zone d'étude.

Premièrement, l'incertitude est limitée par la détermination ou la suppression de zones d'extrapolation (Figure 3.10). Il faut éviter l'extrapolation tant dans l'espace des prédicteurs que dans l'espace géographique (Austin 2007; Merow *et al.*, 2014). L'extrapolation dans l'espace des prédicteurs se produit lorsque les prédictions sont faites dans des zones situées à l'extérieur de la plage des valeurs des prédicteurs utilisées pour l'ajustement du modèle pour les variables prédictives influentes (c'est-à-dire celles qui représentent 95 % de l'influence relative cumulative, voir la section 3.7). L'extrapolation dans l'espace géographique se produit lorsque des prédictions sont faites dans des zones situées à l'extérieur de la zone couverte par les observations, qui peut être représentée par le plus petit polygone convexe autour des observations. Un modèle performant, élaboré à l'aide d'une approche de VC par blocs spatiaux,

convient à l'interpolation (c.-à-d. des prédictions qui ne sont pas extrapolées dans l'espace géographique ou l'espace des prédicteurs). Lorsque les prédictions sont extrapolées, leur performance est inconnue. La pertinence d'un modèle pour l'extrapolation peut dépendre de la méthode de modélisation utilisée. Par exemple, les MLG produisent des relations fonctionnelles prédicteur-réponse qui s'étendent au-delà de la plage des valeurs prédictives dans les données d'apprentissage, tandis que les modèles d'arbres de régression extrapolent dans l'espace des prédicteurs en utilisant une valeur constante – la prédiction à la valeur minimale ou maximale de la plage (Elith et Graham, 2009).

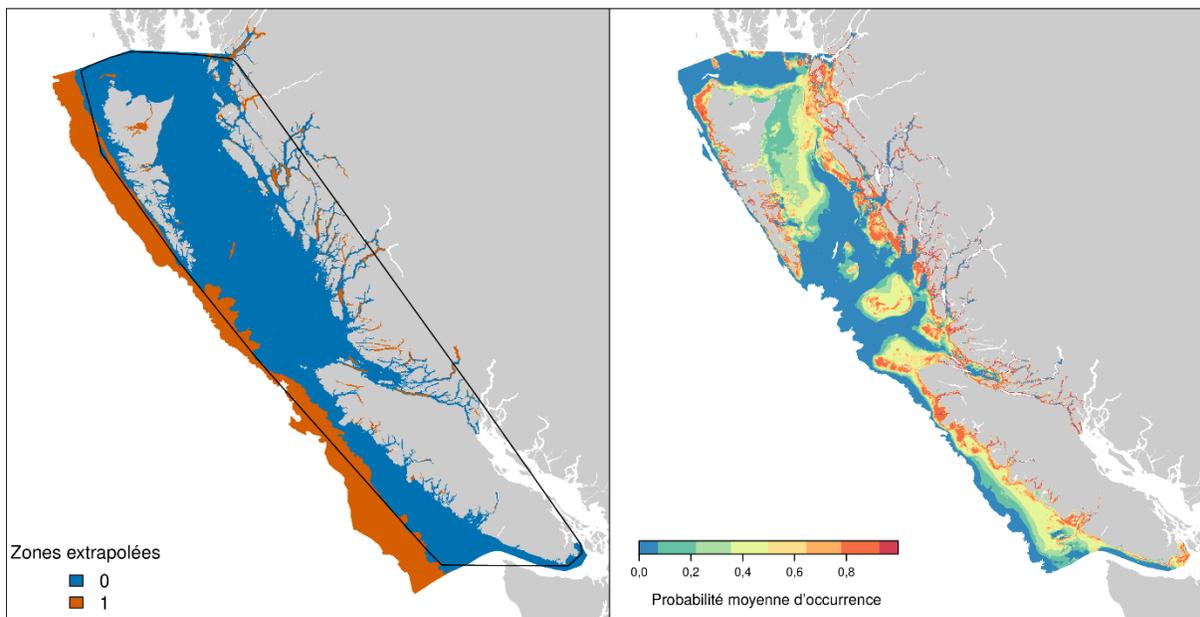


Figure 3.10. Zones extrapolées où les prédicteurs se situaient à l'extérieur de la plage des valeurs utilisées pour l'ajustement du modèle (à gauche). Prédictions de la probabilité d'occurrence moyenne à partir de modèles à validation croisée à cinq plis, en excluant les zones extrapolées dans l'espace des prédicteurs (à droite). La ligne noire représente le plus petit polygone convexe autour des observations.

Deuxièmement, on étudie l'incertitude en comparant les prédictions d'un modèle d'enveloppe fondé sur les connaissances à celles d'un modèle d'ensemble fondé sur les données (Figure 3.11). Les zones où les prédictions du modèle d'enveloppe et du modèle d'ensemble sont très différentes représentent les zones où les prédictions de la probabilité d'occurrence sont les plus incertaines. Sans validation supplémentaire du modèle et vérification sur le terrain, il est impossible de déterminer quelle prédiction est la plus précise dans une zone donnée. Cependant, on peut formuler des hypothèses sur ce qui explique les différences entre notre compréhension actuelle d'une relation espèce-environnement et les relations dérivées de modèles corrélatifs.

Troisièmement, on peut estimer l'incertitude dans le modèle d'ensemble en calculant la variation entre les prédictions du modèle composant individuel (p. ex. celles qui sont obtenues avec différentes méthodes d'ajustement du modèle). Comme la pratique exemplaire recommandée consiste à incorporer plusieurs sources d'erreur dans une seule estimation (Araújo *et al.*, 2019), le cadre intègre à la fois l'incertitude de la VC et l'incertitude méthodologique des modèles composants dans une seule estimation, mais pas l'incertitude des paramètres. Pour estimer l'incertitude méthodologique, on calcule la variance entre les prédictions des différents modèles qui composent l'ensemble (créés à l'aide de différentes méthodes d'ajustement du modèle, p. ex. MLG ou ARS) pour chaque cycle VC et on en fait la

moyenne (Araújo et New, 2007; Jones-Farrand *et al.*, 2011). Pour estimer l'incertitude de la VC, on calcule la variance entre les modèles d'ensemble de chaque cycle VC. Les deux estimations de la variance sont incorporées dans un seul paramètre de l'incertitude en faisant la moyenne des estimations de la variance méthodologique et de la variance de la VC et en calculant la racine carrée pour convertir la variance en écart-type (Figure 3.12). On a choisi l'écart-type plutôt que le coefficient de variation parce qu'il met mieux en évidence les zones de forte incertitude où la probabilité d'occurrence est également élevée et qu'il est plus facile à interpréter. Cette approche de l'estimation de l'incertitude peut s'appliquer à n'importe quelle méthode de modélisation produisant des prédictions à partir de plusieurs modèles. Pour les modèles d'enveloppe fondés sur les connaissances, il faudrait élaborer des prédictions de modèles à partir d'une variété de sources spécialisées ou de connaissances.

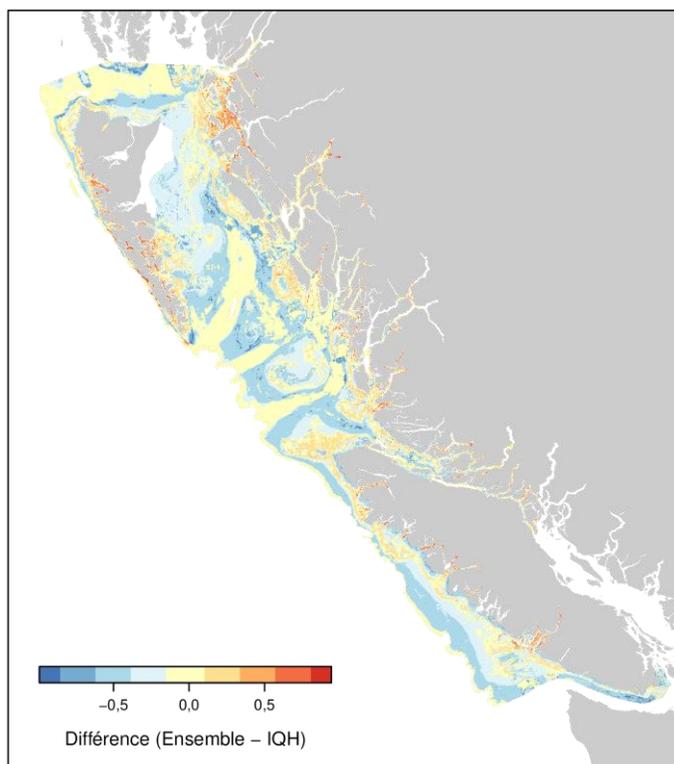


Figure 3.11. Différence entre les prédictions d'un modèle d'enveloppe fondé sur les connaissances et d'un modèle d'ensemble fondé sur les données. Les valeurs positives représentent les zones où le modèle d'ensemble prédisait une plus grande probabilité d'occurrence que le modèle d'enveloppe.

Enfin, on examine l'incertitude en cartographiant les résiduels moyens des modèles à VC pour chaque méthode d'ajustement du modèle (Figure 3.13). On regroupe les résiduels moyens des modèles à VC à une échelle plus grossière (100 fois la résolution des couches matricielles des prédicteurs) et on les cartographie pour visualiser les tendances spatiales qui restent dans les observations après l'ajustement du modèle. Les emplacements où les résiduels sont plus élevés indiquent des zones où les prévisions s'écartent davantage des observations en moyenne et sont donc plus incertaines que les zones où les résiduels sont proches de zéro.

Ces évaluations donnent des estimations qualitatives et quantitatives de la confiance dans les prévisions des modèles ainsi qu'un aperçu des aspects de la qualité des données et du biais d'échantillonnage (en comparant les prévisions des modèles fondés sur les connaissances et celles des modèles axés sur les données). La cartographie des résiduels donne également une idée de la variabilité (c.-à-d. de la non-stationnarité) des relations entre l'espèce et l'habitat.

L'incertitude découlant d'un manque de précision spatiale dans les observations, d'erreurs dans les couches de prédicteurs, de prédicteurs manquants ou d'échelles différentes entre les prédicteurs et les observations n'est pas abordée dans cette évaluation, mais est prise en compte pendant la phase de préparation des données précédant l'application du cadre.

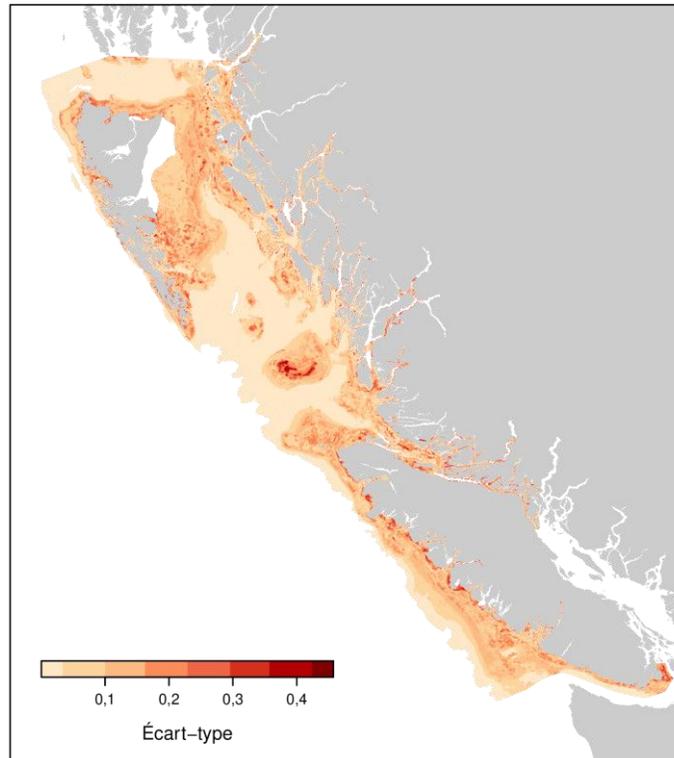


Figure 3.12. Écart-type des prédictions de la probabilité d'occurrence à partir de modèles à validation croisée à cinq plis.

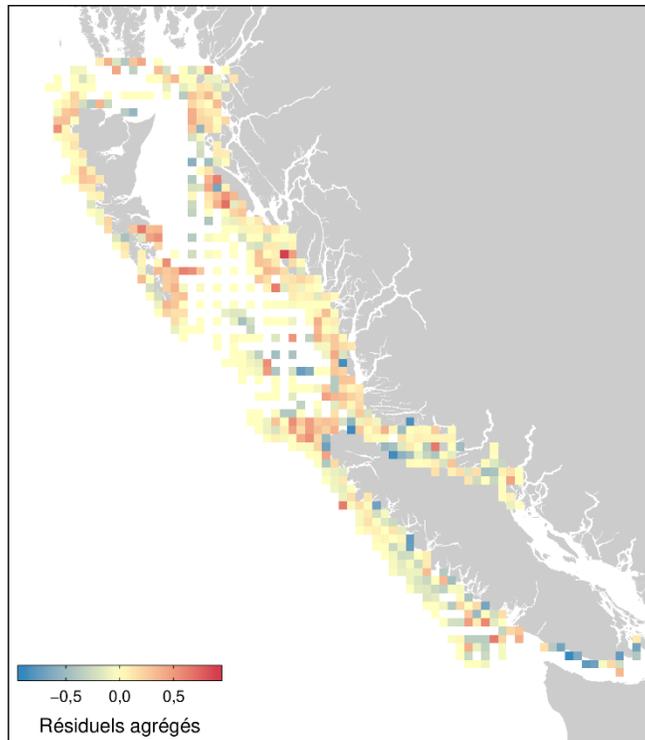


Figure 3.13. Moyenne des résiduels du modèle (agrégés à 10 km) à partir de modèles à validation croisée à cinq plis.

3.7 INTERPRÉTATION DES MODÈLES

Une fois les modèles ajustés, on peut formuler des hypothèses sur les forces relatives et les interactions des nombreux processus qui peuvent influencer sur la répartition des espèces. Ces hypothèses sont fondées sur l'examen de l'influence relative des variables prédictives (Figure 3.14) et des effets marginaux de chaque prédicteur sur les observations (Figure 3.15). Ces graphiques peuvent également aider à formuler une mesure qualitative de la performance du modèle. Par exemple, des relations inattendues dans les effets marginaux (comme une pente positive alors qu'une pente négative était prévue) dénotent des problèmes potentiels de qualité des données (voir des exemples à la section 2) dans les observations ou les prédicteurs. De plus, lorsque l'on utilise des blocs spatiaux, des écarts importants dans l'influence relative ou les effets marginaux entre les modèles à VC peuvent indiquer une non-stationnarité. Il est important de noter qu'il ne s'agit là que de mesures informatives de la stationnarité à l'échelle des prédictions des modèles lorsque les observations sont réparties dans la zone d'étude.

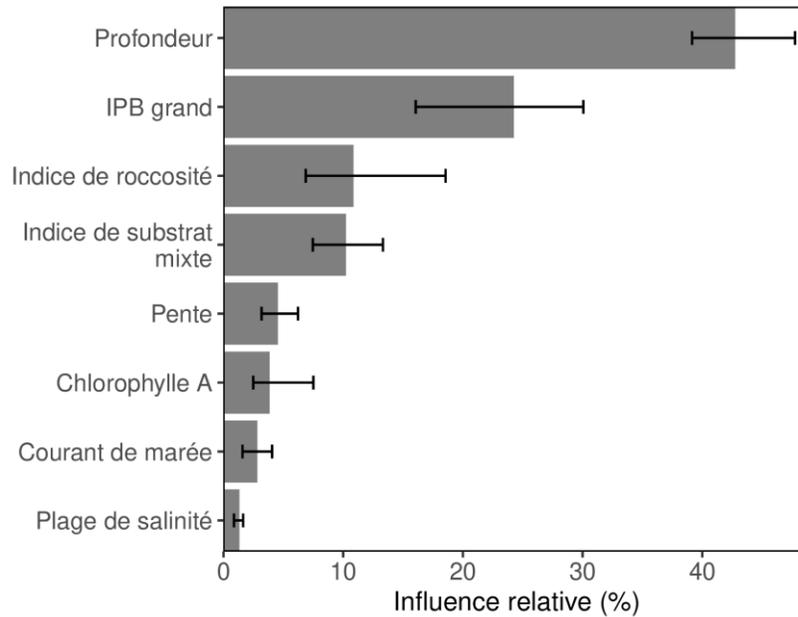


Figure 3.14. Influence relative moyenne des prédicteurs des modèles à validation croisée à cinq plis. Les barres d'erreur représentent l'influence relative minimale et maximale des cinq modèles à VC.

La façon dont on estime l'influence relative des variables prédictives dans un modèle dépend de la méthode de modélisation. Les méthodes de régression utilisent souvent des coefficients normalisés, mais certains ne les trouvent pas fiables (Bini *et al.*, 2009; Hawkins, 2012). Une autre méthode, utilisée dans ce cadre, est la procédure de l'omission d'une variable (« drop-one »), où l'écart résiduel est calculé en omettant chaque prédicteur du modèle un à la fois (Grömping, 2015). L'augmentation de l'écart résiduel lorsqu'un prédicteur est omis est une estimation de son importance relative dans le modèle. Dans le cas des modèles d'apprentissage machine, les méthodes d'estimation de l'influence relative varient. Par exemple, dans le cas des ARS, on estime l'influence relative de chaque prédicteur par le nombre de fois où le prédicteur fractionne une branche, pondéré par la contribution de chaque fractionnement (Friedman et Meulman, 2003).

Les effets marginaux représentent le changement dans les prédictions de la probabilité d'occurrence attribuable à un seul prédicteur. On les estime en faisant varier un seul prédicteur tout en maintenant les autres à leurs valeurs moyennes dérivées des données d'apprentissage utilisées pour ajuster les modèles. Le cadre calcule les effets marginaux pour chaque prédicteur à l'aide de la méthode de la bande d'évaluation d'Elith *et al.* (2005), une approche solide et cohérente pour calculer les effets marginaux, qui convient à toutes les méthodes d'ajustement de modèles.

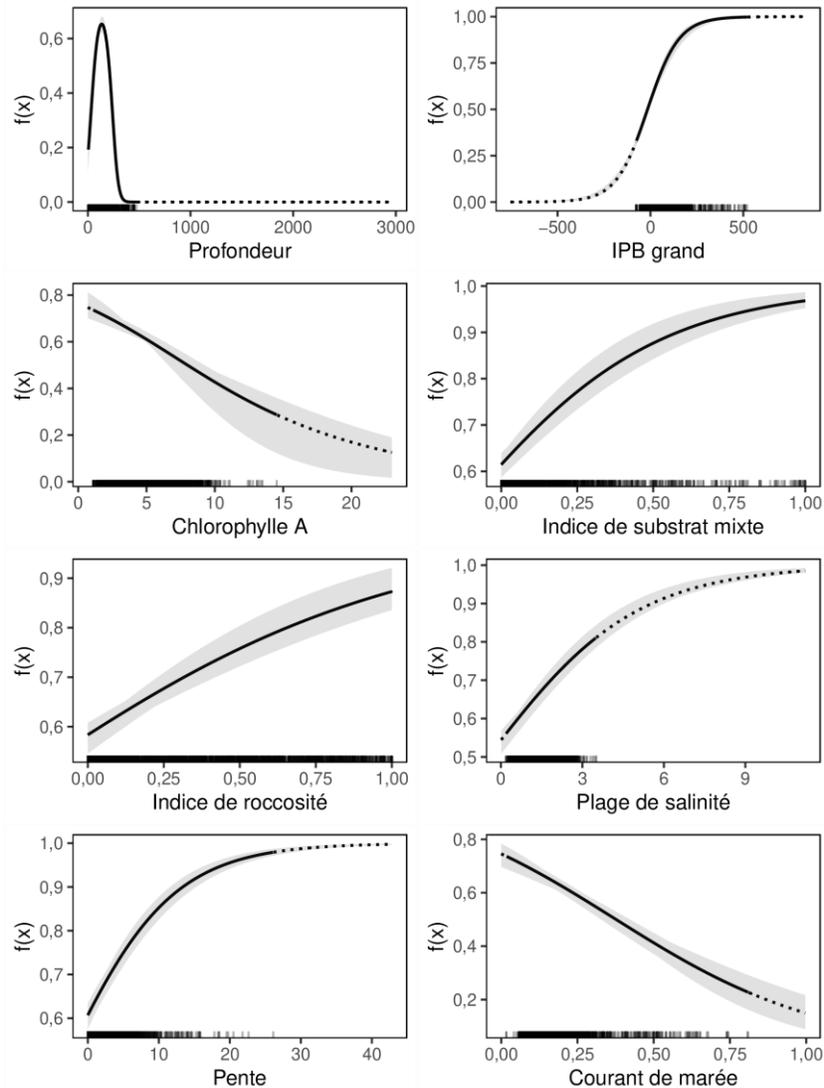


Figure 3.15. Effets marginaux moyens, minimaux et maximaux entre chaque variable prédictive et la réponse prédite des modèles à validation croisée à cinq plis. Les effets marginaux ont été calculés à partir de modèles linéaires généralisés. La ligne pleine noire représente la moyenne, et la zone grisée représente les effets marginaux minimaux et maximaux des cinq modèles. Les effets marginaux sont présentés pour toutes les couches de prédicteurs environnementaux. La ligne pointillée indique une extrapolation dans l'espace des prédicteurs où la variable prédictive se situe en dehors de la plage des valeurs utilisées pour l'ajustement du modèle, ce qui indique qu'il n'y a eu aucune observation de l'espèce dans cet espace. Les marques le long de l'axe des abscisses indiquent la densité des observations dans la plage des valeurs prédictives.

3.8 LE RÔLE DE LA STRUCTURE SPATIALE

Comprendre la tendance de la SAC dans les observations et les résiduels du modèle peut également contribuer à l'interprétation du modèle. Une SAC importante dans les résiduels suggère que les prédicteurs n'expliquaient pas entièrement la structure spatiale des observations (Elith et Leathwick, 2009). Cela peut se produire lorsque des prédicteurs importants manquent, lorsque les prédicteurs disponibles sont mal adaptés aux observations ou lorsque la configuration spatiale est influencée par d'autres facteurs tels que des interactions

biologiques ou des obstacles au déplacement (Bahn et McGill, 2007; Dormann *et al.*, 2007; Elith et Leathwick, 2009; Soberón et Nakamura, 2009).

Le cadre utilise des courbes de variogrammes et la statistique C de Geary pour déceler la SAC positive (Figure 3.16). Les variogrammes illustrent la structure et l'étendue de l'influence de la SAC (Legendre et Fortin, 1989), où l'étendue de la SAC positive est la distance à laquelle la semi-variance atteint une asymptote. Le C de Geary est un paramètre connexe utilisé pour vérifier la signification statistique de la SAC dans un voisinage défini. La statistique varie de 0 à plus de 1, les valeurs inférieures à 1 représentant une SAC positive, et les valeurs proches de 1, une SAC minimale. Le C de Geary est similaire à la statistique I de Moran, souvent utilisée, mais il est moins sensible aux valeurs extrêmes (Legendre et Fortin, 1989).

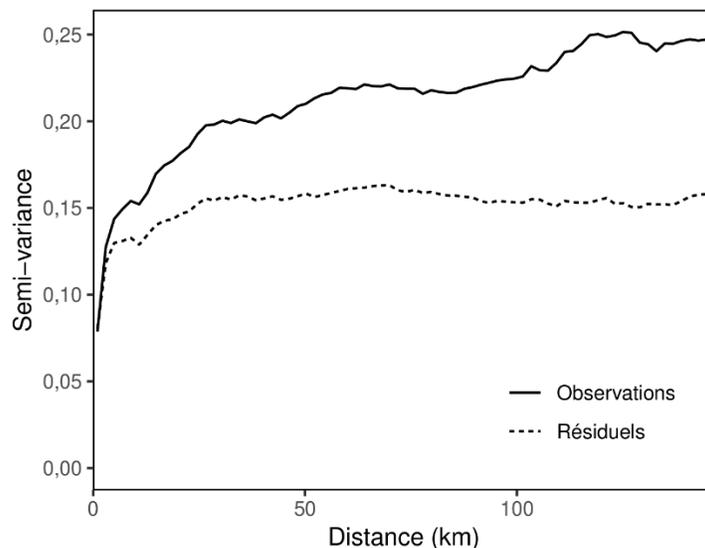


Figure 3.16. Variogramme des observations de la présence-absence des espèces et des résiduels du modèle.

Lorsque les résiduels du modèle conservent leur configuration spatiale, une façon de tenir compte de la variation inexplicée consiste à inclure un terme spatial explicite dans le modèle (Dormann *et al.*, 2007; Merow *et al.*, 2014). L'inclusion d'une mesure de la structure spatiale peut faciliter l'interprétation du modèle en permettant de mieux comprendre l'influence relative des prédicteurs purement spatiaux par rapport aux prédicteurs environnementaux (Hothorn *et al.*, 2011; Keitt *et al.*, 2002; Legendre, 1993; Lichstein *et al.*, 2002). Ce cadre examine la structure spatiale des observations à l'aide de la méthode de l'autocovariable, où un terme prédicteur autocovariable spatial est dérivé *a priori* des observations et inclus dans le modèle (Hughes *et al.*, 2011). Selon Bardos *et al.* (2015), le terme autocovariable est calculé comme la somme des observations proches dans un voisinage prédéfini à pondération symétrique inverse à la distance (Figure 3.17). Dans ce cadre, le voisinage est défini comme la distance maximale du plus proche voisin entre les observations.

On produit des modèles secondaires construits à l'aide d'un terme autocovariable pour faciliter l'interprétation du modèle si l'on trouve une SAC positive importante dans les résiduels des modèles initiaux (voir le flux à la Figure 3.1). L'interprétation de ces modèles à autocovariable peut aider à mieux comprendre en illustrant comment les relations entre les observations et les prédicteurs environnementaux, de même que l'importance relative des prédicteurs, peuvent varier lorsque la structure spatiale locale est prise en compte (Figure 3.18). Le cadre ne recommande pas l'utilisation de modèles à autocovariable pour la prédiction lorsque la

distribution d'échantillonnage des observations ne justifie pas le calcul d'un prédicteur autocovariable exhaustif sur le plan spatial.

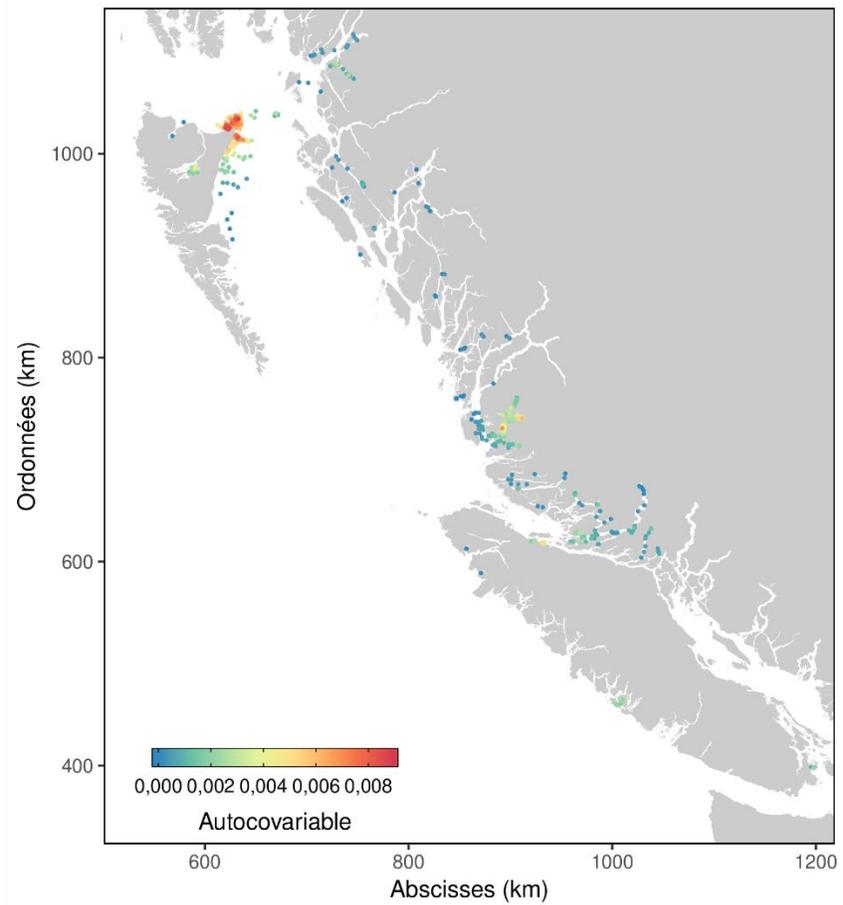


Figure 3.17. Distribution d'un prédicteur autocovariable spatial qui représente la somme pondérée par la distance des observations proches.

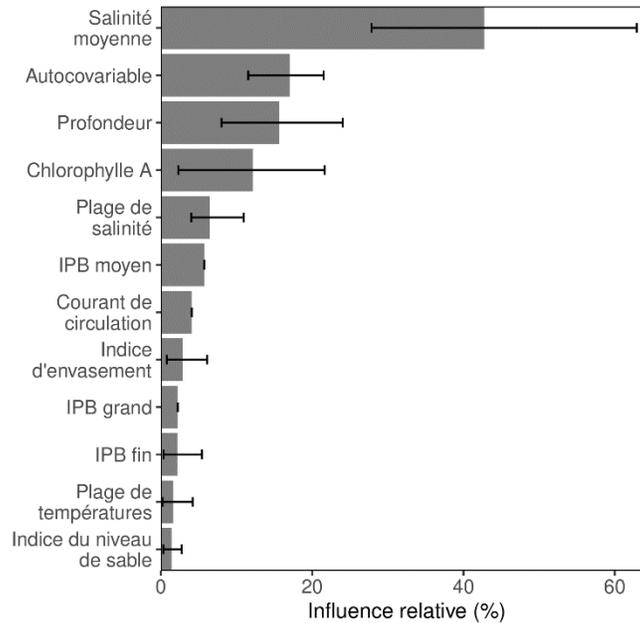


Figure 3.18. Influence relative moyenne des prédicteurs des modèles à validation croisée à cinq plis, y compris l'autocovariable. Les barres d'erreur représentent l'influence relative minimale et maximale des cinq modèles.

4 UNE APPLICATION DU CADRE DE MODÉLISATION DE LA RÉPARTITION DES ESPÈCES

4.1 INTRODUCTION

En 2017, la Garde côtière canadienne a commencé à diriger l'élaboration d'un plan régional d'intervention pour la biorégion du plateau du nord. Ce plan est élaboré en collaboration avec des partenaires fédéraux, provinciaux et des Premières Nations et vise à guider les interventions d'urgence en cas de déversement d'hydrocarbures en mer. À l'appui de ce plan, le Secteur des sciences du MPO a été chargé de combler les lacunes dans les données afin d'aider à déterminer l'emplacement des espèces préoccupantes.

La gestion d'un déversement d'hydrocarbures nécessitant de disposer d'une description bien résolue de la présence des espèces, le cadre de MRE (section 3) a été utilisé pour préparer des modèles d'habitat pour un ensemble diversifié d'espèces dans les biorégions des plateaux du nord et du sud. Afin de maximiser la pertinence pour la gestion, les espèces ont été sélectionnées parmi un groupe plus large d'espèces précédemment désignées comme prioritaires (Gale *et al.*, 2019) en matière de conservation ou très vulnérables au pétrole (Hannah *et al.*, 2017). Des espèces benthiques ont été choisies parce qu'elles sont les plus sensibles aux salissures causées par le pétrole lorsqu'il se dépose. Des espèces ont également été choisies pour représenter une diversité de caractéristiques du cycle vital, d'habitats et de communautés écologiques, avec divers niveaux de disponibilité et de qualité des données, afin d'évaluer la performance du cadre avec différentes combinaisons de modèles écologiques et de données.

Cette application du cadre utilise trois approches de modélisation : les modèles d'indice de qualité de l'habitat (IQH), les modèles linéaires généralisés (MLG) et les arbres de régression stimulée (ARS). Ces méthodes couvrent toute la gamme de la complexité des modèles, et chacune d'entre elles présente des forces et des faiblesses différentes. Les trois méthodes ont

été utilisées pour modéliser la répartition des organismes aquatiques; (IQH: Brooks, 1997; Raleigh *et al.*, 1986); (MLG: Beger et Possingham, 2008; Santoul *et al.*, 2005); (ARS: Leathwick *et al.*, 2006; Leathwick *et al.*, 2008).

Les IQH sont des modèles d'enveloppe simples et fondés sur les connaissances qui dépendent d'un énoncé clair des preuves à l'appui des processus que l'on croit responsables de la création d'un habitat convenable (Brooks, 1997; USFWS, 1981). Ils s'appuient sur les relations publiées et la consultation d'experts, et sont construits manuellement. Bien qu'elles n'exigent pas de données d'observation pour l'élaboration, les pratiques exemplaires recommandent un processus d'étalonnage avec des données sur l'occurrence lorsqu'elles sont disponibles (Brooks, 1997). En l'absence de données sur l'occurrence, les IQH peuvent servir à la fois de référence écologique, de modèle initial auquel comparer les modèles axés sur les données, ou être combinés à des modèles plus complexes dans un ensemble (Jones-Farrand *et al.*, 2011).

Les MLG représentent la classe des méthodes de régression largement utilisées et bien développées dont on se sert pour relier les observations des espèces aux prédicteurs environnementaux. Les MLG exigent que des formes fonctionnelles soient spécifiées, soutiennent des valeurs optimales (à l'aide de fonctions quadratiques) ou des valeurs seuils et peuvent inclure des interactions de manière intuitive. Les MLG ont été choisis plutôt que les modèles additifs généralisés, plus complexes, car ces derniers peuvent s'approcher de la complexité des méthodes d'apprentissage machine, et l'un des objectifs ici était de pouvoir comparer des modèles de complexité différente.

Les arbres de régression stimulée (ARS) sont une puissante méthode d'apprentissage machine basée sur les arborescences pour l'élaboration des MRE. Les arbres de régression divisent récursivement les données de réponse en groupes homogènes, chaque arbre étant divisé en fonction d'une seule variable prédictive (De'Ath et Fabricius, 2000). La stimulation est un processus par étapes qui construit plusieurs arbres à chaque étape pour modéliser les résiduels de l'itération précédente (Elith *et al.*, 2008). Chaque étape utilise une fraction aléatoire des données d'apprentissage (fraction de sous-échantillonnage, « bag.fraction ») pour construire les nouveaux arbres, en améliorant la performance du modèle à chaque étape (Elith *et al.*, 2008). Les interactions sont représentées implicitement, en fonction de la structure des branches des arbres. Cependant, cela peut les rendre plus difficiles à interpréter. Les ARS ont une variété de paramètres qui influencent la complexité du modèle final (p. ex. complexité de l'arbre, taux d'apprentissage, nombre d'arbres de départ, nombre maximal d'arbres à construire et condition d'arrêt).

Chacune de ces méthodes représente une approche différente de l'élaboration du MRE et permet de tirer différents types d'inférences. Cette application du cadre de MRE, faisant appel à différentes approches écologiques, de données et de modèles statistiques, illustre la façon dont le cadre fonctionne dans différents contextes de modèle, fournissant des orientations pour l'élaboration d'autres MRE dans les eaux canadiennes. Les modèles des espèces élaborés dans cette application du cadre de MRE sont destinés aux interventions en cas de déversement d'hydrocarbures et devraient faire l'objet d'un examen plus approfondi (voir Figure 1.1.) pour pouvoir être utilisés dans d'autres applications de gestion.

4.2 MÉTHODES

La présente section décrit l'application de chacune des étapes du cadre (préparation des données, ajustement du modèle, évaluation et prédiction; figure 1.1) aux espèces étudiées. Des détails méthodologiques sont donnés sur la zone d'étude, les espèces modélisées, les données sur l'occurrence et les prédicteurs utilisées, ainsi que les méthodes de modélisation appliquées.

4.2.1 Zones d'étude

Le cadre a été appliqué aux eaux du plateau de la côte canadienne du Pacifique à deux résolutions (figure 4.1). La zone d'étude du plateau (modélisée à une résolution spatiale de 100 m) est délimitée par les coordonnées 47° 57' - 55° 58' N et 123° 6' - 134° 19' O. Elle couvre les biorégions marines des plateaux du nord et du sud (DFO, 2009), la limite occidentale suivant la base de la pente du plateau (à peu près l'isobathe de 2 000 m). La biorégion marine du détroit de Georgie a été exclue parce que les prédicteurs environnementaux du plateau utilisés dans cette analyse ne résolvent pas adéquatement les caractéristiques océanographiques de cette biorégion, et des prédicteurs à plus haute résolution ont été créés spécialement pour cette zone. Des travaux futurs appliqueront ce cadre au détroit de Georgie à l'aide de ces données à plus haute résolution.

La zone d'étude des eaux côtières du nord de la côte centrale (modélisée à une résolution spatiale de 20 m) est située dans la biorégion du plateau du nord et se limite à la région infratidale côtière, de la laisse des hautes eaux à 50 m de profondeur et à 5 km des côtes (Gregn *et al.*, 2013). La zone des eaux côtières s'étend de la passe Portland au nord jusqu'au détroit de la Reine-Charlotte au sud, et couvre plus de 17 500 km de côtes avec de nombreuses îles, chenaux et fjords. Elle a été choisie comme zone d'étude pilote pour ces travaux en raison de la densité plus élevée des données sur l'occurrence des espèces recueillies dans le cadre de projets de développement industriel antérieurs dans la région.

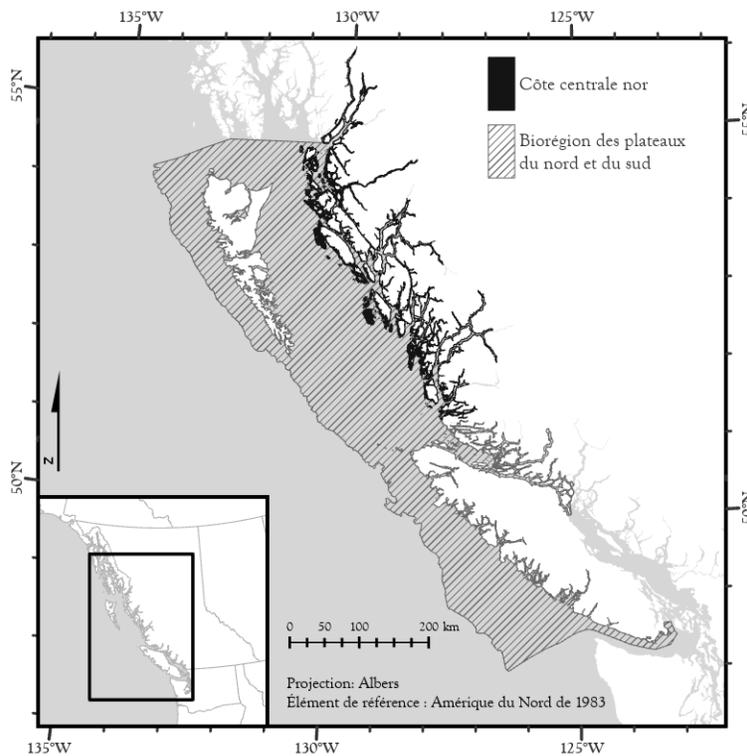


Figure 4.1. Les deux zones d'étude utilisées dans cette analyse. La zone d'étude des eaux côtières du nord de la côte centrale (résolution de 20 m) couvre environ 5 620 km². La zone d'étude du plateau (résolution de 100 m) comprend les biorégions marines des plateaux du nord et du sud et couvre environ 127 840 km².

4.2.2 Prédicteurs environnementaux

Bathymétrie et dérivés

Au total, six couches ont été dérivées de diverses sources de données bathymétriques (Carignan *et al.*, 2013; Davies *et al.*, 2019; Gregr, 2012) aux résolutions spatiales différentes (tableau 4.1). À l'aide d'ArcGIS 10.4, on a rééchantillonné et mosaïqué les couches bathymétriques pour produire des matrices de profondeur pour chacune des deux zones d'étude. Tous les prédicteurs environnementaux ont été projetés dans la projection Albers de la Colombie-Britannique (EPSG:3005).

Les variables de pente, de rugosité et d'indice de position bathymétrique (IPB) ont été dérivées des matrices de profondeur à des résolutions spatiales de 20 m et 100 m. La pente (le taux maximal de changement de profondeur dans chaque cellule) a été calculée en degrés. La rugosité (un indice de rugosité de la surface utilisé comme mesure de la complexité structurale) a été calculée selon Du Preez (2015). Diverses méthodes existent pour calculer la complexité. Nous avons choisi la méthode de rugosité du rapport arc-corde (ACR) (Du Preez, 2015) parce qu'elle découple la rugosité de la pente, ce qui en fait une mesure supérieure de la rugosité de la surface (Du Preez *et al.*, 2016). Les valeurs du IPB représentent la position topographique d'une cellule par rapport à un voisinage de cellules et peuvent être positives (p. ex. haut-fond ou crête) ou négatives (p. ex. fond de vallée). Pour saisir les caractéristiques topographiques locales et à plus grande échelle dans les données, on a créé des couches du IPB à des échelles fines, moyennes et grandes en utilisant différents voisinages (Walbridge *et al.*, 2018). Le IPB fin a été calculé avec une distance de voisinage de 60 à 500 m pour les eaux côtières, et avec une distance de 300 à 2 500 m pour le plateau. Les voisinages des couches du IPB moyen près du rivage et du plateau étaient respectivement de 200 à 2 000 m et de 1 000 à 10 500 m. Les voisinages des couches du grand IPB des eaux côtières et du plateau étaient respectivement de 500 à 5 000 m et de 2 500 à 25 000 m.

Couches océanographiques

En tout, six couches décrivant la température au fond (en degrés Celsius), la salinité (en unité de salinité pratique ou PSU) et la vitesse des courants (marée et circulation en mètres par seconde) ont été dérivées d'un modèle de circulation du Regional Ocean Modelling System (ROMS) de la côte canadienne du Pacifique (tableau 4.1, Masson et Fine, 2012). Le modèle ROMS a une résolution spatiale d'environ trois kilomètres et son domaine s'étend du fleuve Columbia au sud-est de l'Alaska. Les variables de sortie du modèle ROMS ont été tirées d'une simulation rétrospective pour la période allant de 1998 à 2007. La couche du fond était représentée par le plus profond des 30 niveaux sigma, qui suivait la profondeur du fond.

La vitesse du courant a été calculée à partir des vitesses moyennes zonales (u) et méridionales (v), selon une méthode de la moyenne quadratique, après que les vitesses ont été spatialement alignées en les décalant horizontalement grâce à une interpolation linéaire. La température, la salinité et la vitesse du courant de circulation ont été calculées à partir de moyennes sur 15 jours, et la vitesse moyenne mensuelle du courant de marée, à partir de moyennes sur trois heures. La résolution temporelle plus fine des vitesses composantes (3 heures) a permis de résoudre le courant de marée, et les moyennes plus grossières, de 15 jours, ont résolu efficacement la circulation océanique non maréale.

On a ensuite calculé la moyenne temporelle de la température, de la salinité et de la vitesse des courants (marée et circulation) pour la saison printemps/été, d'avril à septembre, sur 10 ans. La saison printemps/été a été choisie parce qu'elle correspond à la période où la majorité des observations des espèces ont été recueillies. Les valeurs minimales et maximales ont été calculées sur 10 ans pour représenter la plage temporelle de la température et de la salinité.

Chlorophylle

Les données sur la concentration moyenne de la chlorophylle a en surface, une approximation de la productivité primaire, ont été dérivées du produit MODIS L2 à une résolution d'un kilomètre. La concentration de chlorophylle a (mg/m^3) obtenue auprès de la NASA a été calculée à partir de la réflectance en utilisant les algorithmes OC4 et CI (pour plus de détails voir Hu et al., 2012). Les données des fauchées quotidiennes entre mars et octobre de 2012 à 2015 ont été téléchargées et mosaïquées par mois. Les mois de novembre à février ont été exclus, car la concentration de nuages est persistante durant cette période. Les données mensuelles ont été interpolées spatialement à l'aide de Spline avec interruptions (Spline with Barriers; ESRI, 2019) avec la laisse des hautes eaux côtières comme barrière pour combler les lacunes qui subsistaient après la mosaïque. Ces lacunes concernaient principalement les eaux côtières et les bras de mer côtiers. La concentration moyenne de chlorophylle a (mg/m^3) a été calculée en faisant la moyenne des mosaïques mensuelles interpolées sur la période de quatre ans.

Tableau 4.1. Sources des variables prédictives environnementales utilisées pour modéliser la qualité de l'habitat de 12 espèces à deux résolutions spatiales (20 m pour les eaux côtières et 100 m pour le plateau). Voir pour la description des unités pour chaque couche de prédicteurs à la section 4.2.2.

Type de données environnementales	Couche(s) de prédicteur	Source(s)	Résolution(s) natives	Années	Zone(s) d'étude	Couches (N)
Bathymétrie	Bathymétrie	MAN bathymétrique Colombie-Britannique en 3 secondes d'arc de la Colombie-Britannique (Carignan <i>et al.</i> , 2013)	3 secondes d'arc	1930-2012	Plateau	6
	Pente					
	Rugosité					
	Grand IPB	MAN à 100 m (Gregr, 2012)	100 m	Plateau		
	IPB moyen	Modèles d'élévation bathymétrique (Davies <i>et al.</i> , 2019)	20 m	Eaux côtières		
	IPB fin					
Océanographie	Salinité moyenne du fond en été	Modèle de circulation régionale de la Colombie-Britannique (Masson et Fine, 2012)	3 km	1998–2007	Eaux côtières	6
	Plage de salinités au fond					
	Température moyenne au fond en été					
	Plage de températures au fond					
	Vitesse moyenne des courants de marée en été					
	Circulation moyenne en été					
Chlorophylle a	Moyenne	Couleur de l'océan de la NASA	1 km	2012-2015	Plateau	1
Fetch	Fetch en été	Script Python (Gregr, 2014)	50 m		Eaux côtières	2
	Fetch minimal					
Substrat	Rocheux	Substrat du fond (Gregr et Haggarty, 2017)	20 m, 100 m		Eaux côtières	4
	Mixte					
	Sablonneux				Plateau	
	Boueux					

Fetch

Le fetch est la distance sur laquelle les vagues poussées par le vent peuvent se former et fournit une mesure approximative de l'exposition à l'action des vagues (Burrows, 2012). Westerbrom et Jattu (2006) affirment que l'exposition aux vagues est un facteur important pour déterminer la répartition des espèces marines le long des côtes rocheuses, et Lessard et Campbell (2007) ont démontré l'importance du fetch pour la répartition des ormeaux. Il était donc essentiel d'inclure une mesure de l'exposition dans les modèles des eaux côtières.

Les valeurs du fetch ont été calculées pour 1 171 638 points (un point par cellule de grille de 50 m) couvrant la zone d'étude des eaux côtières du nord de la côte centrale en utilisant un script Python personnalisé (Gegr, 2014). Pour chaque point, 72 droites de relèvement ont été générées et prolongées jusqu'à 200 km, au maximum, à chaque relèvement de 5 degrés. Les droites ont ensuite été coupées à l'aide de la laisse des hautes eaux côtières. La somme du fetch (une approximation de l'exposition au vent et aux vagues) a ensuite été définie comme la longueur cumulative de la droite de fetch (en m) pour chaque point.

La somme des valeurs du fetch a été convertie en matrices avec des cellules de 50 m. Pour préparer la couche matricielle finale, on a rééchantillonné la couche de 50 m à 20 m selon la méthode bilinéaire, et on l'a alignée avec les autres couches de prédicteurs. Une procédure semblable a été utilisée pour calculer le fetch minimal (une mesure de remplacement de la distance par rapport à la terre), en utilisant la plus petite valeur du fetch (en m) pour chaque point.

Couches de substrat catégoriel et de substrat de voisinage

À la suite des travaux de Gegr et Haggarty (2017), des modèles de substrats catégoriels ont été créés pour les deux zones d'étude. Ces modèles prédisent quatre catégories : roche, mélange, sable et vase. Ils ont été construits selon l'approche de la forêt d'arbres décisionnels, une méthode de classification qui utilise des données d'apprentissage et des couches de prédicteurs (Breiman, 2001). Les données d'observation des substrats proviennent de diverses sources et ont été codées dans les catégories de substrats. Pour préparer les observations en vue de la modélisation, on les a réparties au hasard (tout en préservant la prévalence) en données d'apprentissage et d'essai. La bathymétrie et ses dérivés, ainsi que le fetch et la circulation océanique, ont été utilisés comme prédicteurs. L'ajustement et la prédiction du modèle ont été effectués à l'aide des données d'apprentissage, tandis que les données d'essai retenues (un tiers des observations) ont servi à évaluer le modèle et à produire les statistiques sur la performance. Le code pour l'ajustement du modèle, la prédiction et l'évaluation a été écrit dans le langage statistique R (R Core Team, 2018) et utilisait le progiciel randomForest (Liaw et Wiener, 2002).

Nous avons dérivé des prédicteurs continus du substrat de voisinage à partir des modèles de substrats catégoriels pour refléter l'incertitude dans les modèles catégoriels et mieux caractériser le substrat à son emplacement général plutôt que dans une seule cellule – un attribut qui peut être plus pertinent (c.-à-d. proche) pour des espèces ciblées plus mobiles. L'habitat du sébaste, par exemple, peut être représenté plus efficacement à l'aide d'un indice local de roccosité plutôt que d'une couche catégorielle. Des couches continues d'indices de voisinage ont été créées pour chaque catégorie (rocheux, mixte, sablonneux, boueux) aux deux résolutions spatiales (20 m et 100 m) en appliquant les étapes suivantes. Tout d'abord, une couche binaire a été créée pour chacune des catégories de substrat (roche, mélange, sable, vase). Pour chacune des couches binaires, on a ensuite calculé la somme focale à l'aide d'un voisinage circulaire dont le rayon était de 10 cellules. Enfin, on a converti les couches de la somme focale en couches d'indice en les redimensionnant en $[0, 1]$. Les couches d'indice ainsi obtenues donnent une mesure de la densité du substrat dans le voisinage des cellules.

Prédicteurs spatiaux

Tous les modèles présentés dans cette application ont été produits avec des prédicteurs environnementaux seulement. Les couches de prédicteurs spatiaux (p. ex. autocovariables) étudiées dans le cadre n'ont pas été explorées dans cette application.

4.2.3 Observations des espèces

Sélection des données

Douze espèces benthiques ont été sélectionnées pour cette application. Huit ont été modélisées dans la zone d'étude des eaux côtières (résolution de 20 m) : l'ormeau nordique (*Haliotis kamtschatkana*), la panope du Pacifique (*Panopea generosa*), l'oursin rouge (*Mesocentrotus franciscanus*), la zostère marine (*Zostera* spp.), le varech *Pterygophora* (*Pterygophora californica*), l'étoile ocrée (*Pisaster ochraceus*), le complexe de moules bleues (*Mytilus edulis*, *M. trossulus* et *M. galloprovincialis*) et la palourde du Pacifique (*Leukoma staminea*), et quatre espèces ont été modélisées dans la zone du plateau (résolution de 100 m) : le crabe dormeur (*Metacarcinus magister*), le sébaste à dos épineux (*Sebastes maliger*), le sébaste aux yeux jaunes (*Sebastes ruberrimus*) et la pennatule orangée (*Ptilosarcus gurneyi*).

Pour sélectionner les données sur l'occurrence des espèces, on a tenu compte de la taille de l'échantillon, de l'étendue, du biais d'échantillonnage et de la précision, comme il est indiqué à la section 2.1. La même approche générale a été appliquée à chaque espèce. On a d'abord cherché dans les banques de données du MPO sur les mollusques et crustacés et les poissons de fond des relevés synoptiques ciblés ou systématiques qui échantillonnaient adéquatement les espèces d'intérêt. Pour qu'une espèce soit considérée comme ayant fait l'objet d'un échantillonnage adéquat, il fallait que la méthode de relevé et le type d'engin l'échantillonnent de manière à ce que l'hypothèse d'absence en raison d'un manque de présence soit vraie. Cette détermination reposait d'une part sur les connaissances que possédaient les experts de l'espèce sur la capacité de l'engin de relevé à détecter l'espèce, et d'autre part, sur la prévalence de l'espèce dans les observations du relevé. Une fois qu'un relevé avait été jugé approprié, on a évalué la taille de l'échantillon pour déterminer si des données supplémentaires étaient nécessaires.

C'était le cas lorsque les échantillons des observations de la présence ou de l'absence étaient de petite taille (p. ex. < 100 observations d'absence) ou lorsque la couverture spatiale des observations dans la zone d'étude était limitée dans l'espace géographique (p. ex. aucune observation dans la partie sud) ou dans l'espace des prédicteurs (p. ex. toutes les observations avaient eu lieu sur un habitat rocheux). Bien que la pratique exemplaire consiste à utiliser un seul relevé pour obtenir des données sur l'occurrence des espèces, dans de nombreux cas, un seul relevé n'est pas suffisant pour représenter les occurrences des espèces dans la zone d'étude souhaitée (tableau 4.2). En complétant les données sur l'occurrence des espèces par des relevés supplémentaires, on a tenté de trouver des relevés ayant des taux de détectabilité semblables. Pour toutes les espèces, les données sur l'occurrence proviennent de relevés utilisant le même type d'engin (p. ex. relevés en plongée pour les espèces côtières, palangres pour les sébastes et casiers appâtés pour le crabe dormeur). Même si l'objectif était de maximiser la couverture dans l'ensemble de la zone d'étude et des types d'habitat, toutes les espèces n'ont pas fait l'objet d'un échantillonnage complet dans leur aire de répartition et dans la zone d'étude (p. ex. les observations de crabe dormeur se limitaient principalement aux zones littorales et peu profondes boueuses et sablonneuses, voir la figure A.13).

Une fois qu'un ensemble de données sur l'occurrence d'une espèce avait été jugé suffisant, les données supplémentaires n'ont pas été incluses, même si elles étaient disponibles. Par

exemple, les données du relevé au chalut sur le sébaste étaient disponibles en plus des données du relevé à la palangre. Toutefois, l'inclusion des données sur le chalut introduirait un biais découlant des différences de capturabilité et de zone échantillonnée. De telles différences dans les types d'engins (p. ex. casier ou chalut) influent sur la détectabilité des espèces (p. ex. Wells *et al.*, 2008), ce qui mène effectivement à un biais d'observation. Bien que l'importance de cet effet puisse être mise à l'essai *a priori* et donner une estimation de l'effet du type d'engin, le résultat serait faussé en raison de sa corrélation avec l'habitat (p. ex. les chaluts sont utilisés sur des sédiments plus profonds et plus meubles tandis que les palangres sont utilisées sur des fonds plus durs et moins profonds). Pour dissocier l'effet des engins de pêche de l'effet sur l'habitat, il faudrait procéder à une analyse distincte et avoir un chevauchement spatial important entre les relevés, une situation peu probable étant donné les raisons (c.-à-d. la perte d'engins) pour lesquelles on utilise des engins différents en premier lieu. De plus, pour prédire de nouvelles zones à l'aide d'un modèle de répartition des espèces comprenant un prédicteur du type d'engin, il faudrait avoir une couche associée (et non disponible) décrivant la répartition des types d'engins ou que le modélisateur choisisse le type d'engin le plus approprié pour la prédiction.

Pour quatre des douze espèces sélectionnées (l'étoile ocrée, la moule bleue, la palourde du Pacifique et la pennatule orangée), nous ne disposons pas de données adéquates sur l'occurrence pour réaliser la modélisation dans nos zones d'étude. Les données étaient disponibles pour les quatre espèces, mais elles ont été jugées inappropriées en raison de la petite taille des échantillons (p. ex. < 100 observations de présence), de la faible précision de l'emplacement spatial des observations (p. ex. site plutôt que quadrat) ou de la faible résolution taxonomique (p. ex. observations au niveau des ordres).

Préparation des données spatiales

La précision spatiale des observations des espèces variait d'un relevé à l'autre (p. ex. seulement le lieu du début de l'échantillonnage par rapport aux lieux de début et de fin). Pour décider de la meilleure façon de représenter ces observations dans l'espace, nous avons tenu compte de la précision de l'emplacement spatial des observations par rapport à la résolution des données prédictives. Ainsi, pour tirer le meilleur parti des données prédictives d'une résolution de 20 m, les données sur les espèces côtières devaient contenir des informations spatiales plus précises que celles sur les espèces du plateau.

Toutes les données sur l'occurrence des espèces côtières proviennent de relevés en plongée (tableau 4.2). Tous les relevés en plongée ont suivi un protocole fondé sur des transects perpendiculaires au rivage, à l'exception du relevé sur l'ormeau. Les observations ont été enregistrées pour les quadrats placés le long d'un transect, allant des zones profondes aux zones peu profondes, mais seuls les points de départ et d'arrivée du transect ont été consignés. Pour représenter les espèces côtières à la résolution de 20 m, il fallait connaître les positions des quadrats le long des transects. Pour estimer la position des observations du quadrat, on a créé des points tous les 20 m le long du transect et on a extrait les valeurs de la profondeur de la couche bathymétrique pour chaque point. Les observations du quadrat ont ensuite été assignées aux points qui correspondaient le mieux à la profondeur du quadrat observée (corrigée en fonction de la marée). Si plusieurs quadrats étaient assignés au même point, les observations ont été agrégées. Si la différence entre la profondeur du quadrat enregistrée et la bathymétrie de 20 m était supérieure à 10 m, l'observation était retirée de l'ensemble de données.

Les relevés en plongée du MPO sur l'ormeau suivent un protocole fondé sur une grille (Breen et Adkins, 1979) plutôt que sur un transect. Les observations sont enregistrées par quadrat, les quadrats étant placés selon un quadrillage de 4 x 4 (couvrant une zone de 16 m sur 7 m) dans

un habitat jugé convenable pour l'ormeau. La position de départ de l'échantillonnage est enregistrée. Pour ce relevé, les quadrats individuels ont été agrégés par site et on a utilisé la position de départ pour représenter leur emplacement spatial. Les observations de présence-absence ont été regroupées en assignant la présence lorsqu'on avait au moins une observation de présence dans un quadrat.

Les données sur l'occurrence des espèces dans la zone d'étude du plateau proviennent de relevés effectués à l'aide de palangres et de casiers appâtés (tableau 4.2). Pour les activités de pêche à la palangre, les emplacements spatiaux étaient représentés par des segments de ligne entre les positions de début et de fin de l'événement. Pour les relevés sur le crabe dormeur, l'emplacement spatial des lignes de casiers appâtés était représenté par la position de départ, car il s'agissait de la meilleure information spatiale disponible.

Tableau 4.2. Sommaire des données sur l'occurrence choisies pour modéliser les douze espèces dans cette étude. Une régression linéaire généralisée et des arbres à régression stimulée ont été utilisés pour construire des modèles pour les huit espèces pour lesquelles on n'a pas jugé que les données étaient insuffisantes (DI). Les relevés indiqués avec des données sources (ABL = ormeau; RSU = oursin rouge; BHM = cartographie de l'habitat benthique; cuke = holothurie; HBLL = palangre sur fond dur; IPHC = Commission internationale du flétan du Pacifique) représentent les relevés scientifiques réalisés pour les évaluations ou la surveillance des stocks menées par la région du Pacifique du MPO et ses partenaires industriels.

Espèce	Eaux côtières ou plateau (N/S)	Taille de l'échantillon	Prévalence (%)	Années	Sources de données (présence-absence)	Sources de données (absence seulement)	Type des données spatiales	Précision des données spatiales
Ormeau nordique (<i>Haliotis kamtschatkana</i>)	N	2 293	22	2011-2016	ABL, RSU, BHM	-	Points	Emplacement du site (ABL) ou emplacement estimé le long du transect (RSU, BHM)
Panope du Pacifique (<i>Panopea generosa</i>)	N	9 350	58	2010-2017	GDK, BHM	RSU, Cuke	Points	Emplacement estimé le long du transect
Varech <i>Pterygophora</i> (<i>Pterygophora californica</i>)	N	6 607	3	2010-2017	ABL, Cuke, RSU, BHM	-	Points	Emplacement estimé le long du transect
Oursin rouge (<i>Mesocentrotus franciscanus</i>)	N	3 300	26	2010-2016	RSU, BHM	GDK, Cuke	Points	Emplacement estimé le long du transect
Zostère marine (<i>Zostera</i> spp.)	N	12 567	4	2010-2017	GDK, Cuke, RSU, BHM	-	Points	Emplacement estimé le long du transect
Crabe dormeur (<i>Metacarcinus magister</i>)	S	391	49	1982-2009	Relevés sur le crabe dormeur	-	Points	Position de départ du déploiement de l'engin
Sébaste à dos épineux (<i>Sebastes maliger</i>)	S	4 937	41	2003-2018	HBLL, IPHC	-	Lignes	Position de début et de fin de la palangre
Sébaste aux yeux jaunes (<i>Sebastes ruberrimus</i>)	S	4 937	51	2003-2018	HBLL, IPHC	-	Lignes	Position de début et de fin de la palangre
Pennatule orangée (<i>Ptilosarcus gurneyi</i>)	S	DI	-	-	-	-	-	-
Palourde du Pacifique (<i>Leukoma staminea</i>)	N	DI	-	-	-	-	-	-
Étoile ocrée (<i>Pisaster ochraceus</i>)	N	DI	-	-	-	-	-	-
Complexe de moules bleues (<i>Mytilus edulis</i> , <i>M. trossulus</i> , <i>M. galloprovincialis</i>)	N	DI	-	-	-	-	-	-

Hypothèses des relevés en plongée

Plusieurs hypothèses ont été formulées lors de la sélection des données sur l'occurrence des espèces provenant des relevés en plongée. Dans de nombreux protocoles de relevés en plongée, seules les espèces d'algues dominantes sont consignées dans chaque quadrat, ce qui peut créer de fausses absences. Les experts consultés ont indiqué que, lorsque le varech et la zostère marine sont présents, ils sont probablement dominants. C'est pourquoi nous avons supposé que le manque de présence du varech ou de la zostère marine représentait une véritable absence de ces espèces.

Les espèces non ciblées dans certains protocoles de relevés en plongée ne sont consignées qu'au niveau du site ou du transect, et non au niveau du quadrat. Cependant, nous avons besoin d'observations plus précises (au niveau du quadrat) pour nos modèles des zones côtières. Par conséquent, on a supposé que les absences au niveau du transect indiquaient une absence de cette espèce dans tous les quadrats, mais les présences au niveau du transect n'ont pas été utilisées pour indiquer une présence dans tous les quadrats.

4.2.4 Préparation des données

Les observations des espèces et les données prédictives pour la modélisation ont été préparées selon les méthodes décrites à la section 3.1. Nous soulignons ici les détails supplémentaires qui sont propres à cette application.

La température et la salinité estivales moyennes étaient corrélées, se situant légèrement en deçà du seuil de dix du facteur d'inflation de la variance (VIF). Toutefois, la température et la salinité ont été conservées pour assurer la cohérence entre les modèles d'espèces. Les prédicteurs du substrat n'ont pas tous été inclus pour chaque espèce parce que les prédicteurs sablonneux et rocheux étaient fortement inversement corrélés. Ainsi, pour les espèces dont on sait qu'elles occupent des zones rocheuses, on a utilisé l'indice de roccosité, et les prédicteurs sablonneux et boueux pour les espèces dont on sait qu'elles occupent des sédiments meubles. Le prédicteur du substrat mixte a été utilisé dans tous les modèles d'espèces.

On a appliqué une VC par blocs spatiaux à cinq plis pour construire des modèles MLG et ARS et valider tous les modèles, sauf ceux des espèces pour lesquelles les données sur l'occurrence sont insuffisantes. Pour chaque espèce, on a utilisé des plis identiques de la VC pour la formation et les essais de ces modèles. L'utilisation de données d'essai identiques permet de comparer directement les paramètres de validation des modèles créés selon des méthodes différentes. La taille des blocs spatiaux était fondée sur la fourchette médiane de la SAC pour toutes les couches de prédicteurs environnementaux. Les blocs spatiaux du plateau mesuraient environ 75 km sur 75 km, et ceux des eaux côtières, 35 km sur 35 km.

4.2.5 Modèles d'indice de qualité de l'habitat

Des modèles IQH ont été élaborés pour chacune des douze espèces à partir d'une analyse documentaire et des consultations menées auprès des experts de ces espèces. Contrairement aux modèles corrélatifs (MLG et ARS), les modèles IQH n'utilisaient pas les observations des espèces (décrites à la section 4.2.3) pour établir les relations avec les prédicteurs environnementaux. Plutôt, pour chaque prédicteur environnemental pertinent, on a sélectionné sa contribution à la qualité de l'habitat parmi un ensemble de relations communes (p. ex. Figure 3.4). Chaque relation a ensuite été appliquée au prédicteur environnemental correspondant, ce qui a produit une prédiction univariée de la qualité de l'habitat allant de 0 à 1 (de non convenable à totalement convenable). Les relations entre les prédicteurs ont été élaborées pour les couches de prédicteurs environnementaux décrites à la section 4.2.2. Pour les espèces

(l'oursin rouge) dont on sait qu'elles déplacent leur habitat dans des eaux plus profondes dans les zones plus exposées (en raison de l'action des vagues et du vent), cette interaction a été représentée en divisant les habitats côtiers en zones à faible et à forte exposition à l'aide d'un seuil de 20 (en centaines de kilomètres) et en définissant des courbes de préférence légèrement différentes pour la profondeur dans chacune des deux zones (voir les détails à l'annexe B).

L'ensemble des qualités prédites univariées a ensuite été combiné à l'aide de la méthode des facteurs limitatifs (USFWS, 1981). Cette méthode utilise le prédicteur ayant la qualité la plus basse comme qualité globale pour chaque cellule. Contrairement aux relations compensatoires, une grande qualité dans un prédicteur ne peut compenser la mauvaise qualité dans d'autres. Pour calculer l'influence relative de chaque prédicteur environnemental dans les modèles IQH, on a divisé le nombre de cellules où chaque prédicteur était limitatif par le nombre total de cellules dans la zone d'étude.

L'élaboration des modèles IQH a été un processus itératif, dans lequel des modèles préliminaires ont été préparés à partir des relations écologiques décrites dans la documentation, puis peaufinés par des consultations d'experts. Des documents ont été rédigés pour les consultations d'experts, décrivant les modèles préliminaires IQH et leur justification sur le plan écologique, afin de faciliter le processus de consultation et de maintenir l'uniformité des avis demandés. Les niveaux d'information disponibles variaient d'une espèce à l'autre. Pour certaines espèces, les experts ont tiré parti de leur compréhension générale de la répartition de l'espèce et de ses relations avec les prédicteurs environnementaux. Pour d'autres, les experts disposaient de renseignements plus précis et ils ont affiné les relations entre les prédicteurs environnementaux à l'aide des fourchettes environnementales tirées des données de relevé et vérifié les sorties finales du modèle à partir des zones d'occurrence connues ou des estimations de la densité. Ainsi, les observations des espèces jouaient des rôles variables dans les modèles IQH, certains modèles étant entièrement indépendants des observations des espèces (p. ex. ormeau et zostère marine), tandis que dans d'autres, les observations des espèces appuyaient l'établissement de relations univariées (p. ex. sébaste à dos épineux et panope).

Le processus d'élaboration des modèles IQH différait de celui des modèles axés sur les données (MLG et ARS) puisqu'un seul modèle IQH a été créé pour chaque espèce. Par conséquent, il n'a pas été possible d'estimer la variation des prévisions d'un modèle à l'autre. Une fois les modèles finalisés, huit des douze modèles IQH ont été évalués à l'aide de données d'essai de la VC par blocs spatiaux à cinq plis, ce qui a permis de calculer les moyennes et les écarts-types des paramètres de validation des modèles. On ne disposait d'aucune donnée appropriée pour évaluer les quatre autres modèles (tableau 4.2).

4.2.6 Modèles linéaires généralisés

Cinq MLG, un pour chaque cycle VC, ont été construits pour chaque espèce. Une distribution binomiale des erreurs et une fonction de lien logit ont été utilisées, comme il est recommandé pour les observations binaires (Guisan *et al.*, 2017; Tabachnick *et al.*, 2007). La sélection du meilleur modèle pour chaque ensemble de données de VC a été automatisée grâce à la fonction de torture des données (data dredging) (Barton, 2018). Cette fonction sélectionne les modèles en ajustant toutes les combinaisons possibles de prédicteurs en fonction d'un modèle global et en classant les modèles en fonction d'un critère d'information. Dans ce cas, c'est le critère d'information bayésien (BIC) qui a été utilisé. Le modèle global comprenait tous les prédicteurs de l'ensemble de données prédictives ainsi que des termes quadratiques ou d'interaction jugés importants sur le plan écologique pour une espèce donnée, tels que déterminés par l'information spécialisée tirée des modèles IQH. Par exemple, on sait que

l'ormeau nordique présente une relation quadratique avec l'exposition lorsque la qualité de l'habitat atteint un sommet à des niveaux d'exposition moyens (Lessard et Campbell, 2007). Les termes quadratiques et d'interaction ont été restreints de cette façon en raison des limitations computationnelles de la torture des données avec un ensemble complet de termes d'ordre supérieur et d'interaction. Les meilleurs modèles issus de chacun des cinq cycles VC, ayant chacun une structure de modèle potentiellement différente, ont été évalués à l'aide du pli d'essai, puis utilisés pour prédire la probabilité d'occurrence dans la zone d'étude. La moyenne et l'écart-type des paramètres d'évaluation et des surfaces de prévision ont été calculés pour les cinq modèles.

4.2.7 Arbres de régression stimulée

Cinq modèles ARS, un pour chaque cycle VC, ont été construits pour chaque espèce. Les modèles ARS ont été créés avec des fonctions des progiciels gbm (Greenwell *et al.*, 2018) et dismo (Hijmans *et al.*, 2017) du langage R, en utilisant une distribution de Bernoulli pour la fonction de perte (Ridgeway, 2007). Les modèles ARS exigent un certain nombre de paramètres initiaux qui contrôlent la façon dont le modèle est structuré (tableau 4.3). Les valeurs des paramètres étaient basées sur celles d'Elith *et al.* (2008), mais la fraction de sous-échantillonnage (« bag.fraction ») n'a pas été réglée parce que les essais initiaux ont montré que sa variation n'avait pas d'incidence sur la performance du ARS.

Tableau 4.3. Paramètres initiaux pour les modèles d'arbres de régression stimulée.

Description du paramètre	Valeur initiale
Nombre maximum d'arbres à ajuster	20,000
Nombre d'arbres initiaux	50
Nombre d'arbres par étape	50
Nombre optimal d'arbres	Calculé pendant le réglage
Nombre de plis de validation croisée pour le réglage	5
Tolérance pour la condition d'arrêt pendant le réglage du modèle	0.001
Taux d'apprentissage (contribution de chaque arbre)	0.05, 0.01, 0.005, 0.001
Complexité de l'arbre (nombre de divisions ou de nœuds)	2,4,6,8
Fraction de sous-échantillonnage (proportion des données d'apprentissage sélectionnées à chaque étape)	0.5

Le nombre total d'arbres à construire, le rythme d'apprentissage et la composition de l'arbre ont tous nécessité un réglage pour qu'il soit possible de déterminer leurs valeurs optimales. La procédure de réglage convenait à 16 modèles préliminaires, un pour chaque combinaison unique de taux d'apprentissage et de valeurs de complexité de l'arbre (tableau 4.3), pour chaque cycle VC. Durant la procédure de réglage du modèle, on a divisé les données d'apprentissage à l'aide d'une VC aléatoire à cinq plis en données d'apprentissage interne et en données d'exclusion pour évaluer le critère d'arrêt qui détermine quand le nombre optimal d'arbres a été atteint. L'exclusion d'une partie des données d'apprentissage à utiliser pour la condition d'arrêt, par opposition à l'utilisation du pli d'essai, garde les données d'essai indépendantes de l'ajustement du modèle. La procédure de réglage a construit les modèles de façon itérative, en commençant par 50 arbres et en augmentant de 50 arbres à chaque itération

jusqu'à ce que le maximum de 20 000 arbres ou la condition d'arrêt ait été atteint. Le critère d'arrêt, défini par le paramètre de tolérance, contrôle l'augmentation relative admissible de l'écart d'exclusion par rapport à l'itération précédente avec 50 arbres en moins par rapport à l'itération la plus récente.

Pour sélectionner le meilleur modèle parmi les 16 modèles du réglage, on a estimé les performances prédictives à l'aide de l'écart moyen en pourcentage expliqué à partir des données d'exclusion. Le meilleur modèle – et donc les paramètres optimaux du modèle – a été déterminé pour chaque cycle VC par blocs spatiaux à cinq plis. Selon la règle empirique recommandée par Elith *et al.* (2008), les modèles comptant moins de 1 000 arbres n'ont pas été pris en compte. Par la suite, cinq modèles finaux ont été construits, un pour chacun des ensembles de données de la VC d'apprentissage, en utilisant les paramètres optimaux réglés pour chaque ensemble de données. Les modèles ont été évalués à l'aide du pli d'essai, puis utilisés pour prédire la probabilité d'occurrence dans la zone d'étude. La moyenne et l'écart-type des paramètres d'évaluation et des surfaces de prédiction ont été calculés pour les cinq modèles.

4.2.8 Modèles d'ensemble

Les prédictions des modèles d'ensemble ont été créées en faisant la moyenne des prédictions des modèles MLG et ARS, pondérées en fonction de leur performance (voir la section 3.5). Cinq modèles d'ensemble, un pour chaque cycle VC, ont été construits pour chaque espèce. Il était nécessaire de créer un modèle d'ensemble pour chaque cycle VC pour assurer une validation précise du modèle : si un seul modèle d'ensemble était calculé à partir des prédictions moyennes des modèles MLG et ARS, il ne resterait pas de données d'essai pour évaluer les ensembles. Cette méthode a permis d'évaluer les modèles d'ensemble avec les mêmes données d'essai que les modèles IQH, MLG et ARS. Comme pour les modèles MLG et ARS, la moyenne et l'écart-type des paramètres d'évaluation et des surfaces de prédiction ont été calculés pour les cinq modèles d'ensemble.

4.2.9 Évaluation du modèle

Les modèles ont été évalués à l'aide des paramètres présentés à la section 3.4. Les statistiques dépendantes du seuil étaient fondées sur des seuils calculés en maximisant la TSS à l'aide des données d'essai.

4.3 RÉSULTATS ET DISCUSSION

4.3.1 Performance du modèle par espèce

Les modèles les plus performants étaient ceux du sébaste à dos épineux (AUC d'ensemble = 0,91) et de l'ormeau (AUC d'ensemble = 0,85) (Figure 4.2). La performance des modèles de la panope, du varech, de l'oursin et de la zostère marine était modérément bonne avec des valeurs AUC du modèle d'ensemble supérieures ou égales à 0,8. Les modèles les moins performants étaient ceux du crabe dormeur et du sébaste aux yeux jaunes, bien que les valeurs AUC du modèle d'ensemble aient été adéquates pour les deux, se situant entre 0,71 et 0,78, respectivement. Collectivement, la performance des modèles IQH était médiocre, la majorité ayant donné des valeurs AUC inférieures à 0,7. Le modèle IQH le moins performant était celui de la zostère marine, avec une valeur AUC inférieure à 0,5, ce qui indique que le modèle n'était pas meilleur qu'un modèle aléatoire.

Ces différences de performance ne sont pas uniquement attribuables à la taille de l'échantillon puisque les deux espèces de sébastes avaient des tailles d'échantillon identiques, mais des

paramètres de performance différents. Si la petite taille de l'échantillon a probablement contribué à la mauvaise performance du modèle du crabe dormeur, la distribution limitée des observations dans l'espace géographique et l'espace des prédicteurs a peut-être été un facteur plus important (figure A.13). Les différences de performance peuvent également être liées à des aspects des modèles écologiques sous-jacents, en ce sens que des relations plus solides et plus directes entre les données sur l'occurrence et quelques prédicteurs relativement importants tendent à produire des modèles plus performants. Par exemple, les modèles IQH les plus performants (orveau et sébaste à dos épineux) ont tous deux une forte relation avec un seul prédicteur (voir la Figure 4.3 et la figure A.18, respectivement). Ces prédicteurs satisfont également à l'important critère d'extension au-delà de la plage des valeurs appropriées pour ces espèces. En revanche, bien que le modèle IQH du varech dépende lui aussi fortement d'un seul prédicteur (la profondeur), cette relation est plus distale (la profondeur étant probablement un indicateur de la disponibilité de la lumière).

La comparaison des courbes des effets marginaux et de l'influence relative entre les différentes méthodes du modèle permet de comprendre pourquoi les modèles IQH pour la zostère marine et le crabe ont donné de si mauvais résultats (c.-à-d. les figures A.12 et A.15). Pour la zostère marine, la profondeur et la pente étaient importantes dans les modèles axés sur les données (MLG et ARS), tandis que la profondeur et le substrat étaient les principaux facteurs du modèle IQH (figure A.12). Bien que la profondeur soit importante tant dans les modèles IQH que dans les modèles axés sur les données, la structure de la relation entre la zostère marine et la profondeur diffère. De plus, la pente était notablement absente du modèle IQH; elle n'a pas été incluse parce qu'il n'y avait pas de conseils spécialisés à ce sujet. Cela met en lumière une lacune potentielle du IQH : les connaissances des experts ou décrites dans la documentation peuvent être limitées ou inexistantes pour articuler une relation fonctionnelle. Pour le crabe dormeur, la profondeur et la salinité étaient des facteurs importants dans les modèles IQH et les modèles axés sur les données (figure A.15) et, comme dans l'exemple de la zostère marine, ces prédicteurs différaient de par leur relation fonctionnelle entre les deux modèles. Plus précisément, les modèles axés sur les données montraient une probabilité décroissante d'occurrence à des salinités élevées (> 30) et à des profondeurs plus grandes (> 100), alors que le IQH indiquait que les salinités et les profondeurs au-delà de ces valeurs étaient très appropriées. Ces écarts peuvent être dus à une différence dans l'échelle à laquelle la compréhension et les corrélations écologiques sont développées. Pour le crabe dormeur, les connaissances écologiques proviennent d'observations locales à résolution fine, à l'extrémité inférieure de la plage de salinités, tandis que les prédicteurs de salinité sont relativement grossiers dans l'espace (résolution de 3 km) et dans le temps (moyennes sur 10 ans). Bien que les deux échelles puissent représenter des processus pertinents, les prédicteurs de résolution grossière ne peuvent représenter la compréhension écologique à l'échelle fine, ce qui mène à un décalage d'échelle tel que celui décrit par Wiens (1989).

Ces exemples propres à des espèces illustrent l'utilité diagnostique des paramètres de performance présentés dans ce cadre. Un manque de cohérence dans l'influence relative et les effets marginaux entre les modèles fondés sur les connaissances (p. ex. IQH) et les modèles axés sur les données (p. ex. MLG) fournit des renseignements sur les incertitudes potentielles dans la structure du modèle et la résolution appropriée (spatiale ou temporelle) des prédicteurs environnementaux.

4.3.2 Comparaisons des prédictions des modèles entre les méthodes

Comme illustré ci-dessus, la confiance dans le modèle est accrue par l'uniformité de l'influence relative des prédicteurs et de la forme de leurs effets marginaux entre les différentes méthodes (p. ex. Figure 4.3; voir l'annexe A). Les espèces présentant une telle uniformité sont

probablement bien représentées (c.-à-d. qu'elles font l'objet d'un plus grand réalisme au sens d'Araújo *et al.*, 2019) par les prédicteurs. Les marges d'erreur plus grandes sur les courbes ARS (Figure 4.3), représentant une plus grande différence dans les effets marginaux entre les modèles à VC, constituent une différence notable entre les effets marginaux obtenus à partir des modèles MLG et ARS, malgré leurs formes similaires. Cette tendance est conforme à notre compréhension selon laquelle les modèles ARS s'adaptent mieux aux données d'apprentissage que les MLG et qu'ils auront donc des effets marginaux plus variables pour les modèles construits avec différents ensembles de données d'apprentissage.

Les comparaisons entre les modèles peuvent également aider à évaluer la performance des différentes méthodes. Les modèles ARS correspondent mieux aux données d'apprentissage que les modèles MLG pour toutes les espèces (Figure 4.2), mais la performance des deux modèles était la même lorsqu'on les évalue avec les données d'essai, bien que pour des raisons différentes. Les MLG décrivent des tendances plus générales (relations moins flexibles), et ne relèvent donc pas les tendances à échelle plus fine que les ARS peuvent résoudre. En revanche, les ARS, parce qu'ils sont si flexibles, s'ajustent de trop près aux données d'apprentissage, générant ainsi des tendances moins pertinentes pour les données d'essai. Les deux méthodes n'expliquent donc pas une certaine variation dans les données d'essai.

Cette comparaison de la performance des MLG moins flexibles et des ARS plus complexes confirme les observations mentionnées dans la documentation selon lesquelles les modèles plus complexes (p. ex. ceux qui correspondent aux méthodes d'apprentissage machine) ont tendance à se surajuster aux données d'apprentissage. Toutefois, elle montre également que, lorsqu'on les évalue à l'aide de données spatialement indépendantes (comme celles obtenues avec les blocs spatiaux), la performance mesurée des modèles ARS est conforme à celle des modèles MLG les moins complexes. Cela souligne à la fois l'importance d'une évaluation correcte du modèle et implique que la complexité du modèle, en soi, n'est pas un obstacle à sa transférabilité. Ces résultats concordent avec la documentation (p. ex. Evans *et al.*, 2013; García-Callejas et Araújo, 2016) qui suggère que le surajustement à des données biaisées peut être un problème plus important pour la transférabilité du modèle que la complexité du modèle.

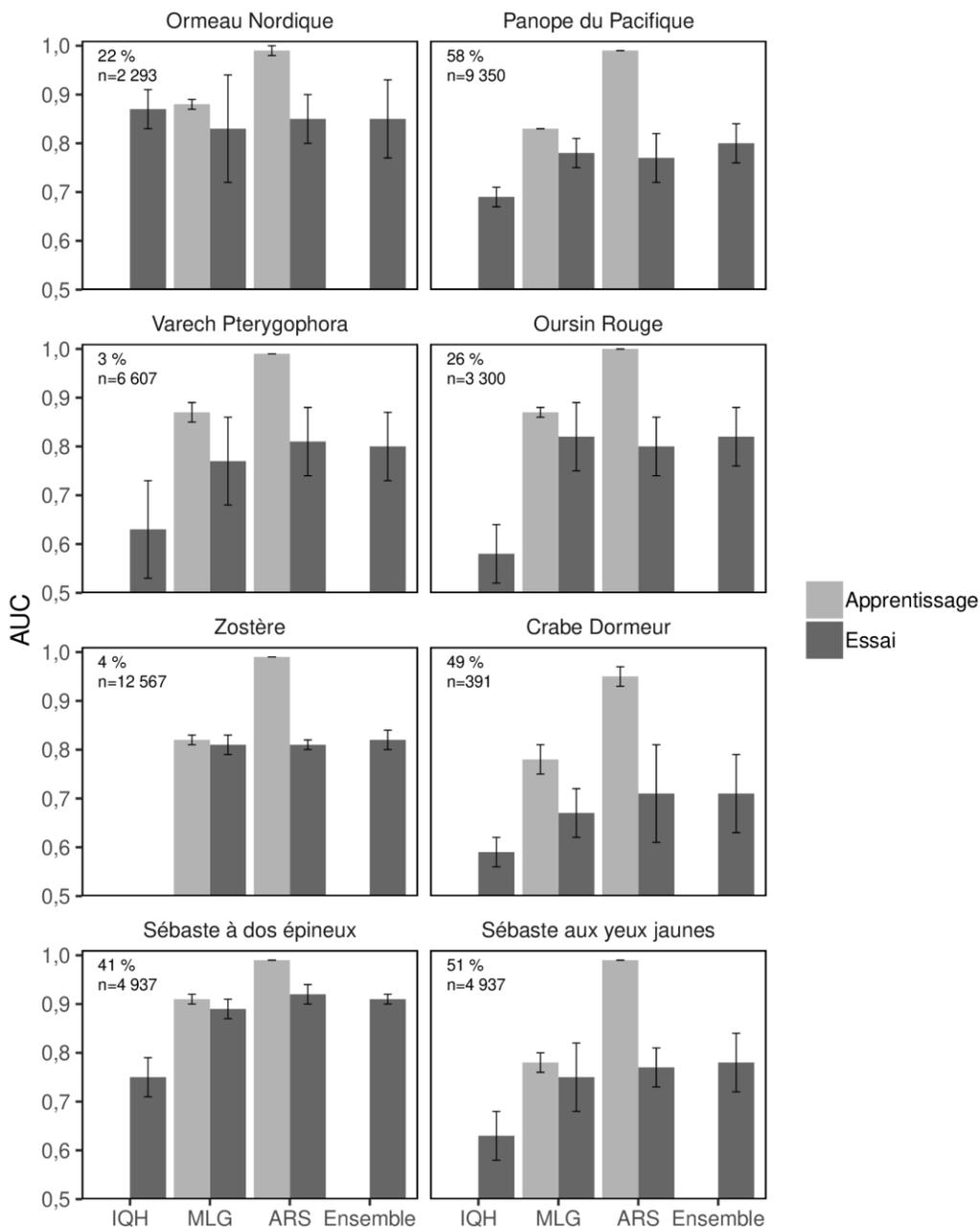


Figure 4.2. Performance des huit modèles de répartition pour les espèces pour lesquelles les données sur l'occurrence sont suffisantes. L'aire moyenne sous la courbe de fonction d'efficacité du récepteur (AUC) est basée sur des essais de validation croisée par blocs spatiaux à cinq plis et est indiquée pour chacun des quatre types de modèles : indice de qualité de l'habitat (IQH), modèle linéaire généralisé (MLG), arbre à régression stimulée (ARS) et ensemble. Les barres d'erreur représentent un écart-type. La prévalence et la taille de l'échantillon sont indiquées dans l'angle supérieur gauche. Les barres non représentées ont une valeur AUC inférieure à 0,5. Les modèles IQH et d'ensemble n'ont pas été évalués à l'aide de données d'apprentissage.

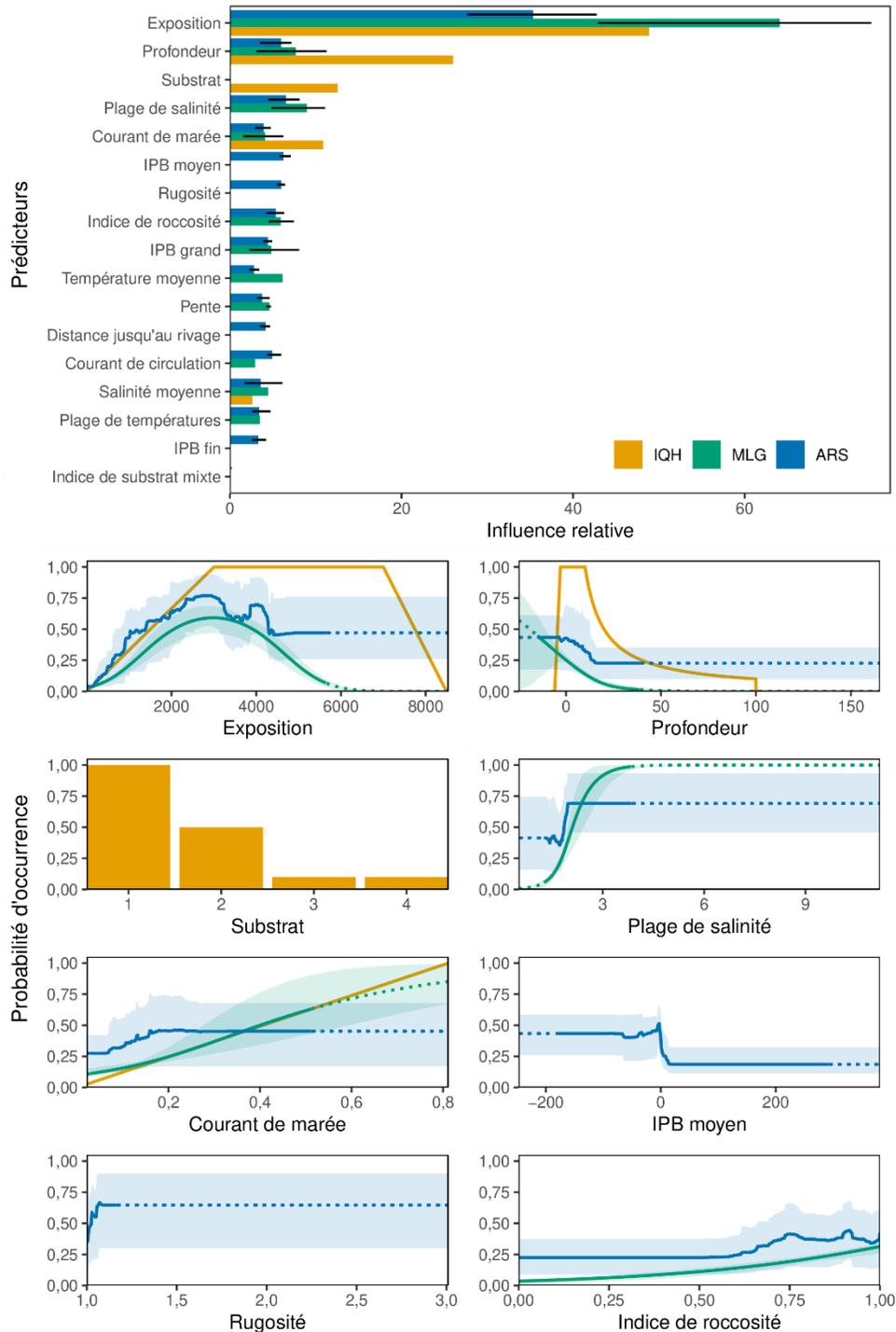


Figure 4.3. Influence relative des prédicteurs et effets marginaux des huit prédicteurs environnementaux les plus influents de la répartition de l'ormeau d'après les modèles IQH, MLG et ARS. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne et les lignes d'erreur montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Sur les courbes des effets marginaux, les lignes pleines noires représentent la moyenne et la zone grisée représente les effets marginaux minimaux et maximaux des cinq modèles à VC à cinq plis. Pour représenter le substrat, on a utilisé une couche de substrat catégoriel pour le modèle IQH et un indice continu de roccosité pour les modèles MLG et ARS.

Dans tous les cas sauf trois (ormeaux, sébaste à dos épineux et varech), la performance des modèles d'ensemble était égale ou supérieure à celle de n'importe lequel des modèles individuels dans le cadre d'une évaluation à l'aide de la VC par blocs spatiaux. Et dans ces trois cas, les modèles d'ensemble se situaient encore bien en deçà d'un écart-type des modèles les plus performants. Bien que construits à l'aide de moyennes pondérées, les résultats des ensembles sont néanmoins conformes aux attentes, à savoir qu'un ensemble à pondération égale sera plus précis qu'au moins la moitié des modèles individuels (Araújo et New, 2007).

Nous avons constaté des tendances semblables dans la distribution des statistiques sur la performance des modèles parmi les modèles MLG, ARS et d'ensemble des huit espèces pour lesquelles des données d'essai étaient disponibles (Figure 4.4). Lorsqu'on les compare d'une espèce à l'autre, les modèles IQH présentaient des paramètres de performance plus variables et obtenaient les résultats les plus faibles pour tous les paramètres, sauf la spécificité (la prédiction correcte des absences), dont la médiane était conforme à celle des autres méthodes de modélisation. Par conséquent, les modèles IQH ont obtenu des résultats semblables à ceux des modèles MLG, ARS et d'ensemble pour prédire l'absence des espèces, mais relativement faibles pour prédire la présence des espèces.

Sur le plan spatial, à grande échelle, la prédiction de la probabilité d'occurrence concorde bien entre les types de modèles (Figure 4.5). Toutefois, un examen attentif révèle des différences évidentes, puisque la variabilité des prédictions augmente avec la complexité du modèle, du modèle IQH au modèle ARS. Cette tendance était évidente chez toutes les espèces.

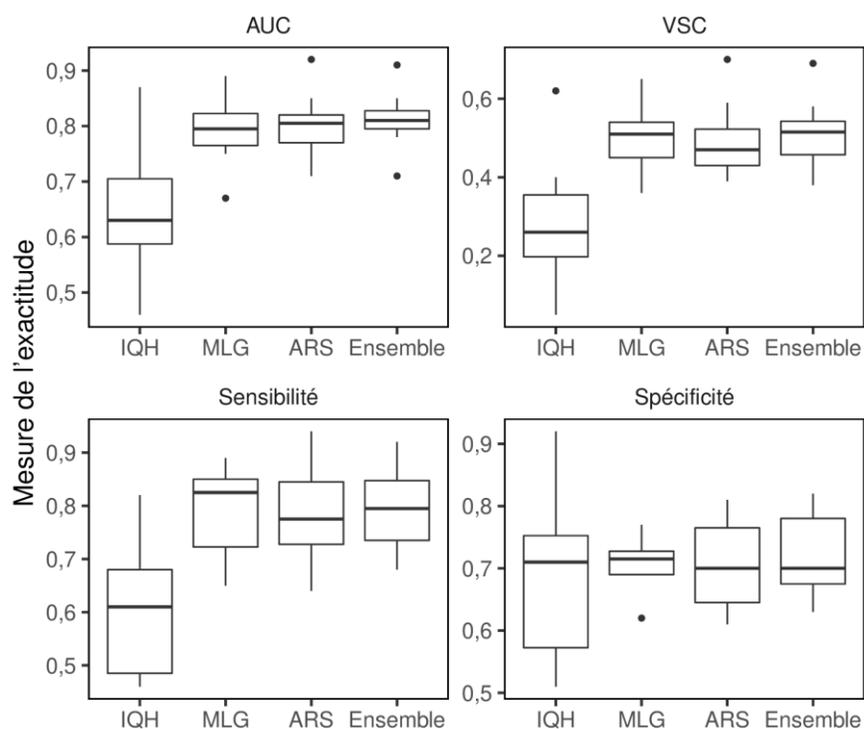


Figure 4.4. Diagrammes montrant la distribution des paramètres de performance des modèles à partir des modèles d'indice de qualité de l'habitat (IQH), des modèles linéaires généralisés (MLG), des modèles d'arbre à régression stimulée (ARS) et des modèles d'ensemble parmi tous les modèles d'espèces mis à l'essai. AUC désigne l'aire sous la courbe de fonction d'efficacité du récepteur, et TSS désigne la « true skill statistic ».

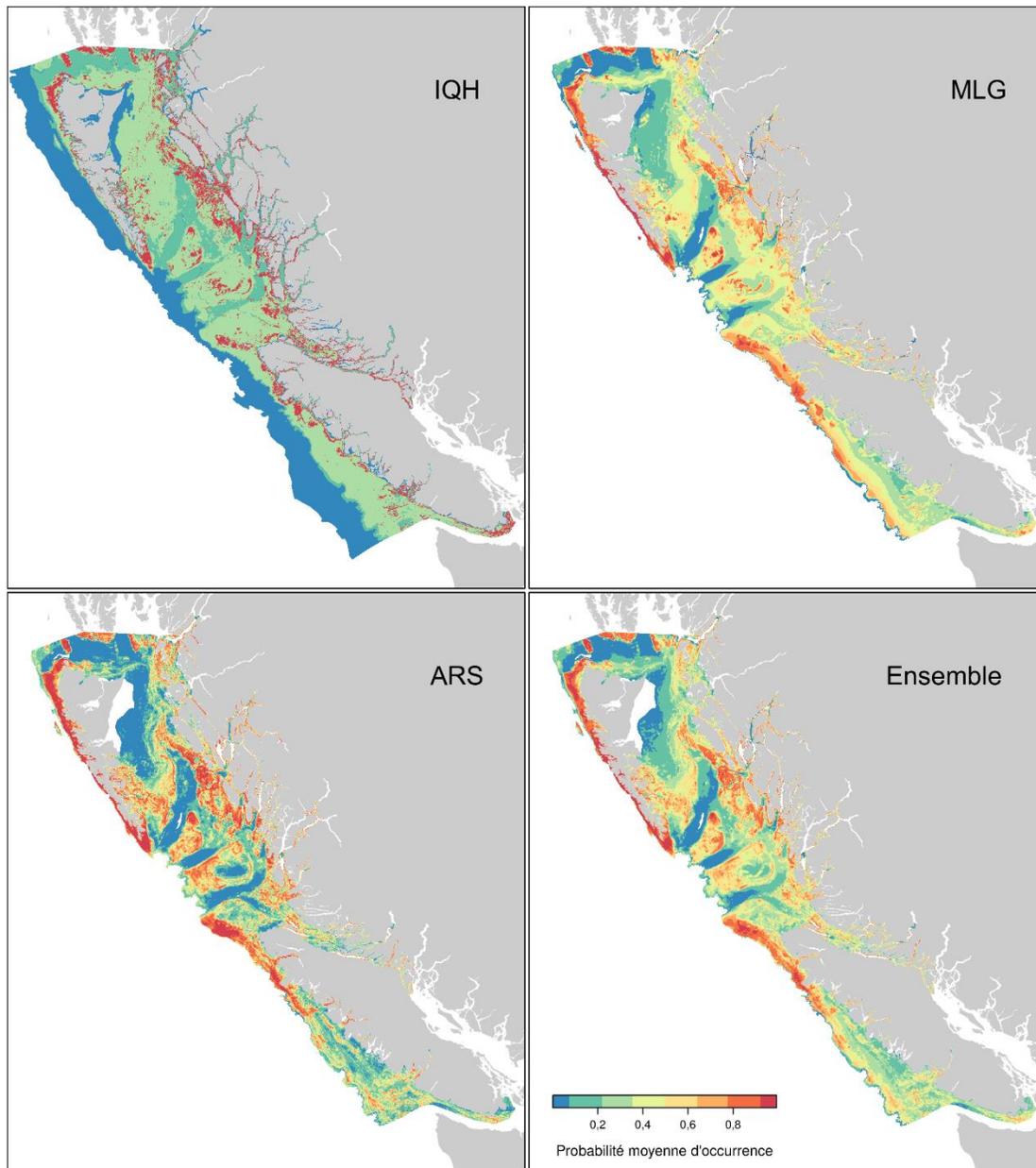


Figure 4.5. Prédications de la probabilité d'occurrence des espèces dans la zone d'étude du plateau d'après le modèle d'indice de qualité de l'habitat (IQH), le modèle linéaire généralisé (MLG), le modèle de l'arbre de régression stimulée (ARS) et le modèle d'ensemble du sébaste à dos épineux. À l'exception de la prédiction du IQH, la probabilité d'occurrence est la prédiction moyenne des modèles à validation croisée à cinq plis. Les prédictions MLG, ARS et d'ensemble sont tronquées pour exclure les zones d'extrapolation dans l'espace des prédicteurs.

4.3.3 Incertitude des modèles

En suivant les méthodes décrites dans le cadre, quatre méthodes ont été utilisées pour examiner l'incertitude du modèle pour chaque espèce avec des modèles d'ensemble. Pour les espèces pour lesquelles seul un modèle IQH a été élaboré, nous n'avons pas été en mesure d'examiner l'incertitude du modèle. Dans cette application, nous présentons l'incertitude du modèle telle qu'elle est mesurée par l'écart-type des prédictions du modèle (voir la section 3.6

et la Figure 3.12) et, à l'annexe A, la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble pour chaque espèce.

Collectivement, les prédictions du modèle ARS présentait une incertitude médiane légèrement plus élevée (écart-type calculé à partir des huit modèles d'espèces pour lesquelles on disposait de données d'essai) que les prédictions du modèle MLG (Figure 4.6). Ce résultat est prévisible, car les modèles ARS sont plus souples et s'adaptent mieux aux données d'apprentissage que les MLG, ce qui entraîne une plus grande variation entre les prédictions des modèles à VC. L'incertitude médiane des prédictions de l'ensemble est nettement plus élevée. Cela s'explique par la mise en commun de la variance des modèles à VC et des modèles MLG et ARS pour intégrer l'incertitude de la VC et l'incertitude méthodologique dans une seule estimation.

L'incertitude globale du modèle d'ensemble, mesurée par l'écart-type moyen des prédictions, variait également selon les modèles d'espèces (Figure 4.7). Les modèles d'espèces dont les valeurs AUC sont plus élevées avaient également tendance à avoir une incertitude moyenne plus faible. Le sébaste à dos épineux, le modèle d'ensemble le plus performant, présentait une incertitude moyenne relativement faible, tandis que le crabe dormeur, le modèle d'ensemble le moins performant, présentait l'incertitude moyenne la plus élevée. L'incertitude moyenne est relativement plus élevée ou plus faible pour certaines espèces compte tenu des statistiques sur leur performance (p. ex. l'ormeau et la zostère marine). L'incertitude peut être très faible pour la zostère marine parce que les prédictions tendent à être faibles en moyenne en raison de l'habitat potentiel relativement limité de l'espèce dans la zone d'étude. Inversement, l'incertitude moyenne peut être plus élevée que la performance du modèle ne le laisse prévoir (p. ex. ormeau) en cas de non-stationnarité dans la zone d'étude. Bien qu'il soit possible de mettre en évidence la non-stationnarité en examinant les paramètres de performance calculés à l'aide des données d'essai créées avec des blocs spatiaux, l'incertitude des prédictions ainsi mesurée peut révéler une variation supplémentaire entre les prédictions du modèle à VC en dehors des zones échantillonnées, qui n'apparaît pas dans les paramètres de performance.

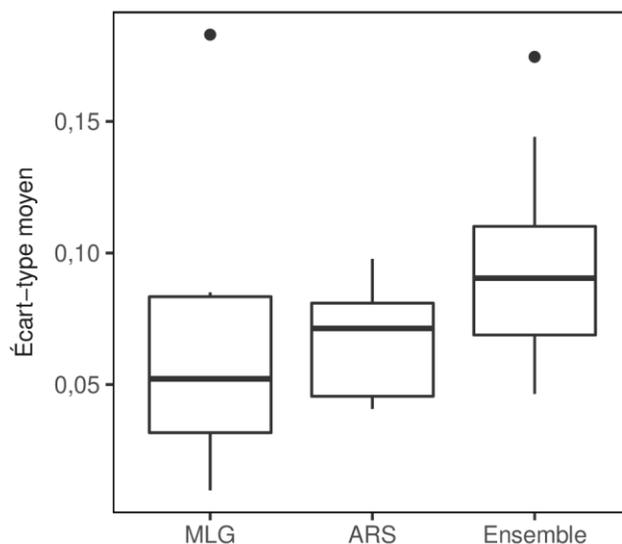


Figure 4.6. Distribution de l'incertitude (en tant qu'écart-type moyen entre les prédictions des espèces) pour trois méthodes de modélisation. L'écart-type moyen des prédictions est basé sur les prédictions des modèles à validation croisée à cinq plis pour toutes les espèces. Les méthodes comprenaient le modèle linéaire généralisé (MLG), le modèle d'arbre de régression stimulée (ARS) et le modèle d'ensemble. L'écart-type des modèles d'ensemble incorpore la variance entre les méthodes MLG et ARS, ainsi que les différents modèles à VC.

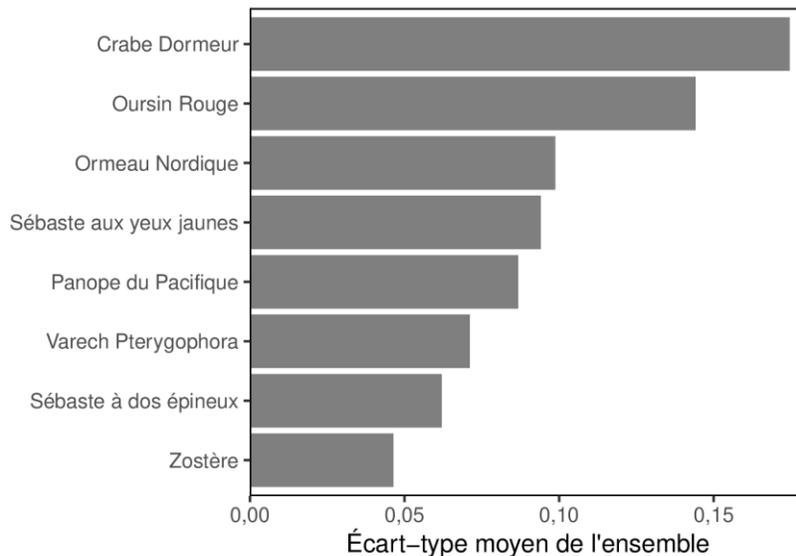


Figure 4.7. Incertitude (en tant qu'écart-type moyen par rapport aux prédictions du modèle d'ensemble) pour les huit espèces pour lesquelles les données d'occurrence sont suffisantes, de la plus élevée à la plus faible.

4.3.4 Comparaison des statistiques sur la performance

Bien que la valeur AUC soit une statistique couramment utilisée, il existe de nombreux documents qui examinent différents paramètres de performance (p. ex. Lawson *et al.*, 2014; Lobo *et al.*, 2008). Il est donc souhaitable de comparer les statistiques sur la performance les plus couramment utilisées dans la documentation afin de déterminer si un paramètre en particulier constitue une mesure plus fiable.

Une comparaison de la façon dont les trois paramètres dépendants du seuil (kappa, TSS et exactitude) changent dans la plage des valeurs seuils d'une espèce (Figure 3.6) montre que le kappa et la TSS se suivent de très près, en restant à moins d'un écart-type l'un de l'autre. En revanche, l'exactitude avait tendance à avoir des valeurs plus élevées et atteignait un plateau à des valeurs seuils plus élevées. Lorsque l'on a comparé les statistiques dépendantes du seuil de tous les modèles d'espèces à la statistique de l'AUC indépendante du seuil, on a dégagé une corrélation presque parfaite (ρ de Spearman = 0,98) entre la TSS et l'AUC (Figure 4.8). En revanche, les mesures de l'exactitude et du kappa présentaient une corrélation plus faible avec l'AUC (ρ de Spearman = 0,77 et 0,60, respectivement). Après inspection, les valeurs aberrantes (sur la gauche du graphique comparant l'AUC et le kappa, Figure 4.8) correspondent aux modèles pour les deux espèces dont la prévalence est la plus faible (varech et zostère marine). Cet effet bien documenté de la prévalence sur le kappa (Allouche *et al.*, 2006; Manel *et al.*, 2001) est également illustré en comparant le kappa à la TSS pour des espèces représentatives à faible et forte prévalence (Figure 4.9). L'effet est frappant lorsque la prévalence est de 5 % et il persiste à des niveaux de prévalence de 25 %, ce qui indique qu'une prévalence proche de 50 % est nécessaire pour rendre le kappa comparable à la TSS et à l'AUC.

Cette comparaison montre clairement comment certaines statistiques sur la performance (AUC, TSS) ont donné des résultats plus uniformes que d'autres (exactitude, kappa).

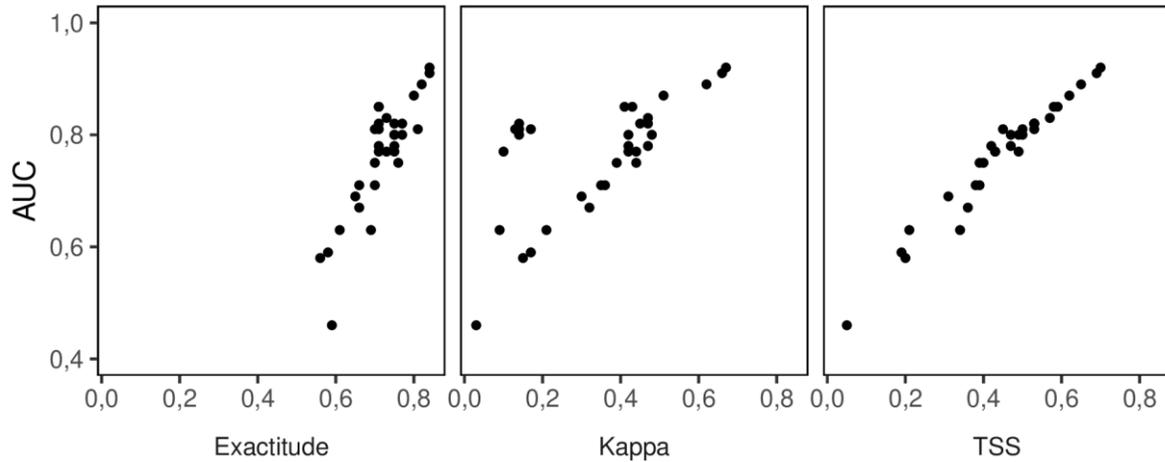


Figure 4.8. Corrélation entre les prédicteurs fondés sur un seuil, exactitude, kappa et TSS (statistique de la qualité vraie) et l'aire sous la courbe de fonction d'efficacité du récepteur (AUC). Toutes les espèces et méthodes de modélisation ont été utilisées pour la comparaison.

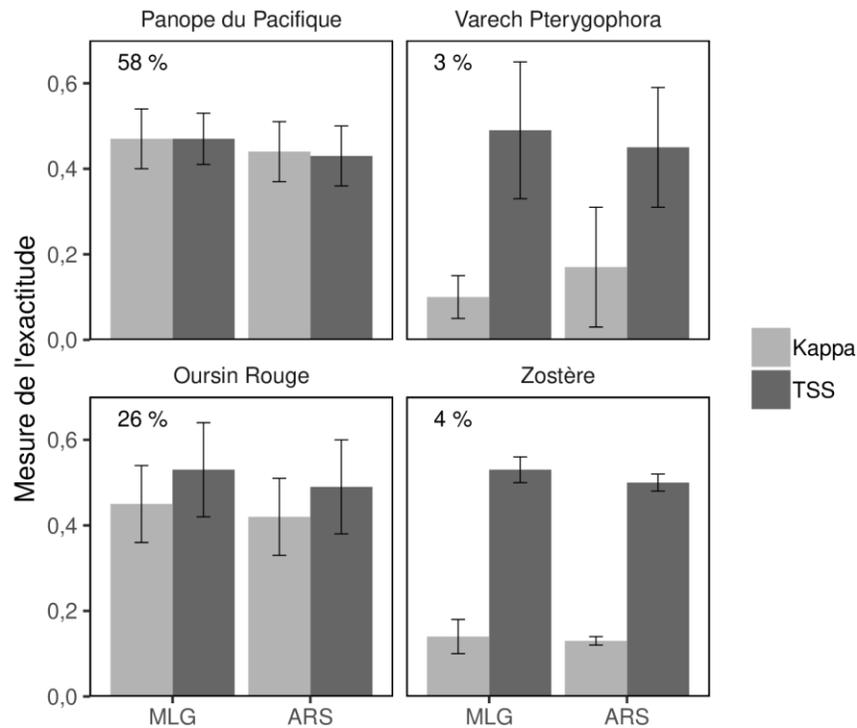


Figure 4.9. Performance moyenne pour le kappa et la TSS (« true skill statistic ») à partir d'une validation croisée à cinq plis, pour des espèces à prévalence élevée et faible. Les paramètres de validation sont présentés séparément pour les modèles linéaires généralisés (MLG) et d'arbre de régression stimulée (ARS). La prévalence est indiquée dans l'angle supérieur gauche pour chaque espèce. Les barres d'erreur représentent un écart-type.

4.4 CONCLUSIONS

Les douze modèles de répartition des espèces présentés ici montrent comment le cadre décrit dans la section 3 peut être appliqué efficacement pour créer des prédictions crédibles. Les prédictions de la probabilité d'occurrence ainsi obtenues serviront à étayer les interventions d'urgence en cas de déversement d'hydrocarbures en mer. Pour les huit espèces pour lesquelles les observations de l'occurrence sont adéquates, les paramètres de performance du modèle et les mesures de l'incertitude spatialement explicites fournissent aux gestionnaires des outils pour évaluer la confiance dans les prédictions du modèle dans un site d'intérêt donné. Pour les quatre espèces pour lesquelles on a jugé que les données sont insuffisantes, les modèles d'enveloppe préliminaires fournissent une base de référence à partir de laquelle on pourra élaborer des modèles de répartition fondés sur les données une fois que les données d'observation seront disponibles. En suivant un processus normalisé d'élaboration de modèles, on a produit les modèles de répartition efficacement tout en assurant la cohérence entre les espèces. Il était essentiel d'utiliser un ensemble unique de données prédictives (c.-à-d. un modèle de données cohérent) pour que le cadre puisse traiter efficacement plusieurs modèles d'espèces. De plus, au cours du processus, le cadre (ainsi que les scripts R connexes) a facilité les mises à jour lorsque de nouvelles données sur l'occurrence ou de meilleurs prédicteurs sont devenus disponibles. L'approche itérative appliquée pour réévaluer la pertinence des données d'entrée du modèle pendant la phase d'interprétation du modèle et pour peaufiner par la suite les modèles (voir la Figure 1.1.) a permis d'améliorer la performance du modèle dans de nombreux cas et constitue un élément précieux du cadre.

Cette application a également montré l'utilité d'une approche à plusieurs modèles pour élaborer des MRE. L'élaboration de modèles d'enveloppe fondés sur les connaissances a aidé à définir la compréhension écologique actuelle des douze espèces et a servi de point de référence auquel les prédictions du modèle d'ensemble axé sur les données ont pu être comparées. L'élaboration des prédictions du modèle d'ensemble a permis d'estimer l'incertitude en examinant la variation entre les modèles composants. L'influence relative et les effets marginaux des prédicteurs (voir l'annexe A) ont également contribué à guider les évaluations de la fiabilité des modèles des espèces individuelles.

Cette application a donné des résultats particulièrement bons pour des espèces d'intérêt commercial ou de conservation, pour lesquelles on disposait d'un vaste échantillon d'observations appropriées, même si les données d'observation pour certaines espèces (p. ex. le crabe dormeur) étaient spatialement biaisées vers un habitat plus approprié. Pour ces espèces, des observations supplémentaires obtenues grâce à un programme d'échantillonnage bien conçu permettraient d'améliorer les prédictions du modèle. Dans le cas des espèces pour lesquelles on dispose de données d'observation de qualité (c.-à-d. des données sur l'abondance à une résolution appropriée et étendues avec un biais d'échantillonnage spatial limité) et qui présentent une bonne association avec les variables prédictives élaborées ici, les modèles de densité des espèces sont probablement possibles et fourniraient des renseignements supplémentaires pour les interventions en cas de déversement d'hydrocarbures et d'autres besoins de gestion.

5 RECOMMANDATIONS

Les recommandations suivantes ont été formulées au cours de l'élaboration du cadre et de son application.

- Suivre la pratique de modélisation établie en définissant clairement les objectifs du modèle. Ils guident l'élaboration des données et des modèles statistiques et sont essentiels à

l'interprétation des résultats des modèles, en partie parce qu'ils permettent d'établir des distinctions claires entre ce qui est techniquement erroné (p. ex. un prédicteur mal mis à l'échelle) et ce qui sort de la portée (p. ex. les étapes du cycle biologique des juvéniles), ce qui limite la mauvaise interprétation des modèles.

- Créer des modèles d'enveloppe fondés sur les connaissances, indépendamment de la disponibilité des données d'observation. De tels modèles clarifient la compréhension écologique actuelle de l'espèce, aident à déterminer les prédicteurs essentiels à la caractérisation efficace de l'habitat et fournissent un moyen de calculer l'incertitude du modèle en comparant les prédictions des modèles fondés sur les connaissances à celles des modèles axés sur les données.
- Intégrer les CEL autant que possible afin de compléter les connaissances écologiques scientifiques. Ces connaissances seront particulièrement utiles dans les situations où les données sont rares, et elles pourraient être incluses dans les modèles d'enveloppe et combinées à titre de connaissances préalables avec des approches axées sur les données dans un cadre bayésien.
- Examiner la cohérence spatiale et temporelle de la résolution entre le prédicteur et les données sur l'occurrence des espèces afin d'évaluer la pertinence du modèle de données et de cerner les écarts d'échelle avant l'ajustement du modèle.
- Tenir compte de l'influence de la qualité des données d'occurrence sur les mesures de la performance du modèle. Si la confiance dans la qualité des données sur les espèces est faible, des scores de performance élevés ne seront pas à eux seuls indicatifs d'un bon modèle, car ces mesures reflètent également la qualité des données sur l'occurrence (c.-à-d. que si les données d'entrée sont biaisées, l'évaluation le sera aussi). Différents ensembles de données d'observation peuvent produire des modèles dont les mesures de performance sont très semblables, mais dont les prédictions de la probabilité d'occurrence sont très différentes dans la zone d'étude.
- Envisager de réduire la taille de l'échantillon si une partie des données sur l'occurrence (p. ex. un relevé ou une méthode d'échantillonnage en particulier) nuit à la qualité globale des données. Il faut souvent trouver un compromis entre la taille de l'échantillon et la qualité des données, surtout si les données sur l'occurrence sont compilées à partir de sources multiples. Un nombre moins grand de données de qualité peut produire un meilleur modèle qu'un nombre plus grand de données de piètre qualité.
- Utiliser la VC par blocs spatiaux pour construire des modèles et évaluer la performance. La création de blocs pour les données d'observation réduit la dépendance entre les données d'essai et les données d'apprentissage et produit des scores d'évaluation du modèle qui sont une mesure plus précise de la transférabilité du modèle. Pour développer la perspective la plus large sur la performance du modèle, calculer des paramètres à partir des données d'apprentissage et d'essai pour évaluer l'ajustement du modèle et la performance prédictive.
- Examiner les graphiques de validation tels que ceux inclus dans ce cadre (c.-à-d. les graphiques de seuil, la matrice de contingence graphique et la courbe ROC) en plus des statistiques individuelles pour évaluer la performance du modèle. Les représentations graphiques fournissent plus d'informations que les nombres individuels et sont donc plus informatives, surtout pour comparer plusieurs modèles.

-
- Les statistiques AUC, indépendantes du seuil, et TSS, dépendantes du seuil, sont recommandées plutôt que les statistiques du kappa, sensibles à la prévalence, et les statistiques d'exactitude, moins descriptives.
 - S'il reste de l'autocorrélation spatiale dans les résiduels après l'ajustement du modèle, envisager un prédicteur spatial tel qu'un terme autocovariable pour faciliter l'interprétation et la compréhension du modèle. Les modèles à prédicteurs spatiaux ne conviennent pas toujours à la prédiction, mais leur comparaison à des modèles uniquement environnementaux révèle l'importance relative des variables environnementales une fois la structure spatiale prise en compte, et peut aider à déterminer les processus manquants en estimant une structure spatiale inexplicite.
 - Diagnostiquer les erreurs possibles dans les données et les modèles statistiques en recherchant les relations inattendues dans les effets marginaux et l'influence relative des prédicteurs. De tels diagrammes peuvent donner un aperçu important de l'incertitude structurelle du modèle et du biais non pris en compte dans les données sur l'occurrence.
 - Créer des prédictions d'ensemble à partir de plusieurs modèles (p. ex. à partir de différentes méthodes d'ajustement des modèles) et calculer les variations entre les modèles sources pour obtenir une mesure spatialement explicite de l'incertitude des modèles.
 - Envisager d'utiliser des données recueillies de façon indépendante pour une validation supplémentaire du modèle dans la mesure du possible.

6 GLOSSAIRE

Cette section contient un ensemble de termes dont la définition est essentielle à la compréhension du contenu du présent document. Nous avons inclus ce glossaire pour fournir aux lecteurs une référence centralisée à la terminologie technique.

AUC : Aire sous la courbe de fonction d'efficacité du récepteur. Paramètre courant servant à estimer la performance des modèles de qualité de l'habitat.

AIC : Le critère d'information d'Akaike est une mesure largement utilisée de la performance relative des modèles, visant à équilibrer le compromis entre l'ajustement des modèles aux données et la parcimonie. Pour une méthode corrélative donnée et un ensemble d'observations et de données prédictives, on l'utilise pour choisir, parmi un ensemble fini de modèles, celui dont l'AIC est le plus faible. Pour les modèles statistiques, ce critère est une fonction des degrés de liberté dans les données d'observation.

Biais : Effet de l'échantillonnage non aléatoire d'une variable. Des données biaisées peuvent obscurcir la tendance d'intérêt et conduire à des inférences incorrectes.

BIC : Semblable à l'AIC, le critère d'information bayésien est également utilisé pour sélectionner des modèles de rechange. Fondé en partie sur la fonction de vraisemblance, il applique une pénalité un peu plus importante que l'AIC et, par conséquent, est plus susceptible de choisir des modèles moins complexes.

Coefficient de variation : Rapport de l'écart-type sur la moyenne.

Complexité : La gestion de la complexité est un élément clé de la construction de modèles. Elle a été décrite de diverses façons dans la documentation. Ici, sauf indication contraire, elle fait référence à la complexité de la structure du modèle et équivaut aux degrés de liberté utilisés par le modèle.

Validation croisée (VC) : Processus de partitionnement des données en plis multiples (K) et d'utilisation répétée de K-1 plis comme plis de données d'apprentissage pour l'ajustement du modèle, et d'utilisation du pli restant pour l'évaluation.

Écart : Mesure de la qualité de l'ajustement calculée à partir des résiduels d'un modèle statistique.

Modèles d'écosystèmes : Vaste catégorie de modèles utilisés pour représenter un aspect particulier d'un système écologique, habituellement en termes mathématiques.

Modèle d'indice de qualité de l'habitat (IQH) : Modèle d'enveloppe fondé sur les connaissances, qui s'appuie sur la documentation et la consultation d'experts pour décrire les relations entre l'espèce et l'environnement.

Échantillons indépendants : Un échantillon (ou collecte) de données est indépendant d'un autre s'il n'y a pas de relation entre eux qui permettrait de tirer des conclusions sur l'un à partir de l'autre.

Connaissances écologiques locales (CEL) : Connaissances écologiques sur un lieu, détenues par un groupe précis de personnes, portant sur les organismes vivants et leur relation avec l'environnement, y compris les connaissances écologiques traditionnelles transmises de génération en génération.

Étendues du modèle : Espace géographique sur lequel le modèle est construit.

Modèle de présence-absence : Ces modèles reposent sur des données d'absence relativement certaines, obtenues par une surveillance ou des visites répétées. La fiabilité des données d'absence dépend des caractéristiques de l'espèce (p. ex. biologie, comportement, histoire), de son abondance locale et de sa facilité de détection, ainsi que du plan du relevé (Hirzel *et al.*, 2006).

Modèle de présence seulement : Modèle qui utilise uniquement les observations de présence, ce qui rend difficile l'évaluation de la spécificité du modèle. Ce problème a été résolu de deux façons : 1) l'utilisation de pseudo-absences pour permettre l'utilisation de cadres de présence-absence; 2) l'évaluation des prédictions en fonction de leur écart par rapport au modèle aléatoire (Hirzel *et al.*, 2006).

Prévalence : Proportion des valeurs de présence ou d'abondance dans un ensemble de données dépendantes (les données d'absence doivent être incluses dans les données).

Sensibilité : Proportion d'observations de présence qui sont correctement prédites en tant que présence.

Autocorrélation spatiale (SAC) : Les observations proches dans l'espace sont plus susceptibles d'être similaires que celles qui sont plus éloignées.

Modèle de répartition des espèces (MRE) : Modèle prédictif de la probabilité d'occurrence des espèces dans l'espace. Bien qu'elles soient généralement fondées sur des corrélations entre les observations des espèces et les caractéristiques de l'habitat, les interactions des espèces sont de plus en plus prises en compte.

Spécificité : Proportion des observations d'absence qui sont correctement prédites en tant qu'absence.

Stationnarité : Hypothèse clé sur les processus qui sous-tendent les relations entre les espèces et leur habitat. Ces processus sont dits stationnaires lorsqu'ils ne varient pas dans la zone d'étude, dans l'espace et dans le temps. La vraisemblance que cette hypothèse soit fautive augmente avec l'étendue du modèle.

Seuil : Seuil utilisé pour convertir les prédictions continues (probabilité d'occurrence) en prédictions binaires (présence/absence).

Variance : Mesure de la dispersion d'un ensemble de valeurs autour de leur moyenne.

7 REMERCIEMENTS

Les auteurs tiennent à remercier les nombreuses personnes qui ont contribué à la réussite de ce travail. Anders Knudby, Chris Rooper et Javier Murillo-Perez ont fourni des examens réfléchis et bien informés qui ont permis d'apporter d'importantes améliorations à ce manuscrit. Matthew Grinnell a apporté une contribution significative aux scripts R qui construisent et évaluent les modèles axés sur les données. Leslie Barton, Michelle Bigg, James Boutillier, Dominique Bureau, Dan Curtis, Sarah Davies, Kieran Forge, Dana Haggarty, Christine Hansen, Elise Keppel, Joanne Lessard, Dan Leus, Tammy Norgard et Emily Rubidge ont fait part de leurs connaissances écologiques sur lesquelles reposent les modèles IQH. Sean Anderson, Katherine Bannar-Martin, Mauricio Carrasquilla, Lisa Lacko, Shannon Obradovich, Maria Surry, Katie Gale et Kim Conway ont fourni des conseils et du soutien au sujet des données sur les espèces et des données environnementales. Enfin, les auteurs remercient chaleureusement tous ceux qui ont contribué à la collecte et à la gestion des nombreuses sources de données sur les espèces et l'environnement utilisées ici, sans lesquelles ce travail n'aurait pas été possible.

8 RÉFÉRENCES

- Allouche, O., Tsoar, A., and Kadmon, R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *J. Appl. Ecol.* **43**: 1223-1232. doi:10.1111/j.1365-2664.2006.01214.x.
- Anderson, O.F., Guinotte, J.M., Rowden, A.A., Tracey, D.M., Mackay, K.A., and Clark, M.R. 2016. Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. *Deep-Sea Research Part I: Oceanographic Research Papers* **115**: 265-292. doi:10.1016/j.dsr.2016.07.006.
- Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* **22**(1): 42-47.
- Araújo, M.B., and Peterson, A.T. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology* **93**(7): 1527-1539. Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., and Naimi, B. 2019. Standards for distribution models in biodiversity assessments. *Science advances* **5**(1): eaat4858.
- Augustin, A.N.H., Muggleston, M.A., and Buckland, S.T. 1996. An Autologistic Model for the Spatial Distribution of Wildlife. *J. Appl. Ecol.* **33**: 339-347.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecol. Model.* **200**(1): 1-19.
- Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol. Model.* **157**: 101-118.
- Bahn, V., and McGill, B.J. 2007. Can niche-based distribution models outperform spatial interpolation? *Global Ecol. Biogeogr.* **16**: 733-742. doi:10.1111/j.1466-8238.2007.00331.x.
- Bahn, V., and McGill, B.J. 2013. Testing the predictive performance of distribution models. *Oikos* **122**(3): 321-331. doi:10.1111/j.1600-0706.2012.00299.x.
- Bardos, D.C., Guillera-Aroita, G., and Wintle, B.A. 2015. Valid auto-models for spatially autocorrelated occupancy and abundance data. *Methods in Ecology and Evolution* **6**: 1137-1149. doi:10.1111/2041-210X.12402.
- Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. *J. Appl. Ecol.* **43**(3): 413-423.
- Barton, K. 2018. MuMin: Multi-model inference. R package version 1.15.6.; 2016.
- Beger, M., and Possingham, H.P. 2008. Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. *Mar. Ecol. Prog. Ser.* **361**: 1-13.
- Belisle, A.C., Asselin, H., LeBlanc, P., and Gauthier, S. 2018. Local knowledge in ecological modeling. *Ecol. Soc.* **23**(2): 14.
- Bell, D.M., and Schlaepfer, D.R. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. *Ecol. Model.* **330**: 50-59.
- Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F., Akre, T.S., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araujo, M.B., Baselga, A., and Beck, J. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. *Ecography* **32**(2): 193-204.

-
- Breen, P., and Adkins, B. 1979. A survey of abalone populations on the east coast of the Queen Charlotte Islands, August 1978. *Fish. Mar. Serv. Manusc. Rep* **1490**: 125.
- Breiman, L. 2001. Random forests. *Machine learning* **45**(1): 5-32.
- Brooks, R.P. 1997. Improving Habitat Suitability Index Models. *Wildlife Society Bulletin (1973-2006)* **25**(1): 163-167.
- Burnham, K.P., and Anderson, D.R. 2004. Model selection and multi-model inference: a Practical Information-Theoretic Approach. Second ed. Springer-Verlag, New York.
- Burrows, M.T. 2012. Influences of wave fetch, tidal flow and ocean colour on subtidal rocky communities. *Mar. Ecol. Prog. Ser.* **445**: 193-207.
- Carignan, K., Eakins, B., Love, M., Sutherland, M., and McLean, S. 2013. Bathymetric Digital Elevation Model of British Columbia, Canada: Procedures, Data Sources, and Analysis. NOAA National Geophysical Data Center (NGDC). 8 pp.
- Carrasquilla-Henao, M., Yamanaka, K.L., Haggarty, D., and Juanes, F. 2018. Predicting important rockfish (*Sebastes* spp.) habitat from large-scale longline surveys for southern British Columbia, Canada *Can. J. Fish. Aquat. Sci.* **13**: 1-13. doi:10.1139/cjfas-2017-0458.
- Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. *Fish Fish.* **10**: 235-251.
- Colding, J., and Folke, C. 2001. Social taboos: “invisible” systems of local resource management and biological conservation. **11**(2): 584-600.
- Dallas, T.A., and Hastings, A. 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. *Global Ecol. Biogeogr.* **27**(12): 1448-1456.
- Davies, S.C., Gregr, E.J., Lessard, J., Bartier, P., and Wills, P. 2019. Bathymetric elevation models for ecological analyses in Pacific Canadian coastal waters. *Can. Tech. Rep. Fish. Aquat. Sci.* 3321: vi + 38 p.
- De’Ath, G., and Fabricius, K.E. 2000. Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. *Ecology* **81**: 3178-3192.
- Dormann, C.F., Purschke, O., Márquez, J.R.G., Lautenbach, S., and Schröder, B. 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. *Ecology* **89**(12): 3371-3386.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., and Leitão, P.J. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**(1): 27-46.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D.W., Peres-Neto, P.R., Ohlemüller, R., Reineking, B., Kühn, I., Schröder, B., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* **30**: 609-628. doi:10.1111/j.2007.0906-7590.05171.x.
- Du Preez, C. 2015. A new arc–chord ratio (ACR) rugosity index for quantifying three-dimensional landscape structural complexity. *Landscape Ecol.* **30**(1): 181-192.
- Du Preez, C., Curtis, J.M., and Clarke, M.E. 2016. The structure and distribution of benthic communities on a shallow seamount (Cobb Seamount, Northeast Pacific Ocean). *PloS one* **11**(10): e0165513.
-

-
- Eger, A.M., Curtis, J.M., Fortin, M.-J., Côté, I.M., and Guichard, F. 2016. Transferability and scalability of species distribution models: a test with sedentary marine invertebrates. *Can. J. Fish. Aquat. Sci.* **74**(5): 766-778.
- Elith, J., and Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32**(1): 66-77.
- Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol., Evol. Syst.* **40**: 677-697.
- Elith, J., Burgman, M.A., and Regan, H. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. *Ecol. Model.* **157**: 313-329.
- Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**(4): 802-813.
- Elith, J., Ferrier, S., Huettmann, F., and Leathwick, J. 2005. The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. *Ecol. Model.* **186**(280-289).
- Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., and Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Divers. Distrib.* **17**(1): 43-57.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S., and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129-151.
- Ericksen, P., and Woodley, E. 2005. Using multiple knowledge systems: benefits and challenges. In *Ecosystems and human well-being: multiscale assessments*. Island Press, Washington, DC, USA: 85-117.
- ESRI. 2019. ArcGIS. Environmental Systems Research Institute, Redlands, CA.
- Evans, M.R., Grimm, V., Johst, K., Knuuttila, T., De Langhe, R., Lessells, C.M., Merz, M., O'Malley, M.A., Orzack, S.H., and Weisberg, M. 2013. Do simple models lead to generality in ecology? *Trends Ecol. Evol.* **28**(10): 578-583.
- Foster, S.D., Dunstan, P.K., Althaus, F., and Williams, A. 2015. The cumulative effect of trawl fishing on a multispecies fish assemblage in south-eastern Australia. *J. Appl. Ecol.* **52**(1): 129-139.
- Fourcade, Y., Besnard, A.G., and Secondi, J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. *Global Ecol. Biogeogr.* **27**(2): 245-256.
- Freeman, E.A., and Moisen, G.G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecol. Model.* **217**(1-2): 48-58.
- Friedman, J.H., and Meulman, J.J. 2003. Multiple additive regression trees with application in epidemiology. *Stat. Med.* **22**: 1365-1381. doi:10.1002/sim.1501.
- Gale, K.S.P., Frid, A., Lee, L., McCarthy, J., Robb, C., Rubidge, E., Steele, J., et Curtis, J.M.R. 2019. Cadre d'identification des priorités en matière de conservation écologique pour la planification d'un réseau d'aires marines protégées et son application dans la biorégion du plateau nord. *Secr. can. de consult. sci. du MPO. Doc. de rech.* 2018/055. viii + 204 p.
-

-
- Galipaud, M., Gillingham, M.A.F., David, M., and Dechaume-Moncharmont, F.X. 2014. Ecologists overestimate the importance of predictor variables in model averaging: A plea for cautious interpretations. *Methods in Ecology and Evolution* **5**: 983-991. doi:10.1111/2041-210X.12251.
- García-Callejas, D., and Araújo, M.B. 2016. The effects of model and data complexity on predictions from species distributions models. *Ecol. Model.* **326**: 4-12.
- Garrison, B., and Lupo, T. 2002. Accuracy of bird range maps based on habitat maps and habitat relationships models. In *Predicting species occurrences. Issues of scale and accuracy*. Edited by J. Scott and P. Heglund and M. Morisson and J. Haufler and M. Raphael and W. Wall and F. Samson. Island Press, Washington. pp. 367-375.
- Greenwell, B., Boehmke, B., and Cunningham, J. 2018. gbm: Generalized Boosted Regression Models. R package version 2.1.4. .
- Gregr, E.J. 2011. Insights into North Pacific right whale *Eubalaena japonica* habitat from historic whaling records. *Endangered Species Research* **15**: 223-239.
- Gregr, E.J. 2012. BC_EEZ_100m: A 100 m raster of the Canadian Pacific Exclusive Economic Zone. SciTech Environmental Consulting, Vancouver BC.
- Gregr, E.J. 2014. Fetch Geometry Calculator Version 1.0 – User Guide. SciTech Environmental Consulting, Vancouver, BC. 5 pp.
- Gregr, E.J., and Chan, K.M.A. 2014. Leaps of Faith: How implicit assumptions compromise the utility of ecosystem models for decision-making. *Bioscience* **65**(1): 43-54.
- Gregr, E.J., and Haggarty, D. 2017. Background substrate and the integration of nearshore and deep water classifications (Draft). SciTech Environmental Consulting, Vancouver, BC. 26 pp.
- Gregr, E.J., Lessard, J., and Harper, J. 2013. A spatial framework for representing nearshore ecosystems. *Prog. Oceanogr.* **115**: 189-201.
- Gregr, E.J., Palacios, D.M., Thompson, A., and Chan, K.M. 2018. Why less complexity produces better forecasts: An independent data evaluation of kelp habitat models. *Ecography* **42**(3):428-43. doi:10.1111/ecog.03470.
- Grimm, V., Augusiak, J., Focks, A., Franke, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P., and Railsback, S.F. 2014. Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecol. Model.* **280**: 129-139.
- Grömping, U. 2015. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software* **17**. doi:10.18637/jss.v017.i01.
- Guinotte, J.M., and Davies, A.J. 2014. Predicted deep-sea coral habitat suitability for the U.S. West Coast. *PLoS ONE* **9**. doi:10.1371/journal.pone.0093918.
- Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* **135**: 147-186.
- Guisan A, Thomas C. Edwards J, Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* **157**:89-100.
- Guisan, A., Thuiller, W., and Zimmermann, N.E. 2017. *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.
-

-
- Hannah, L., St. Germain, C., Jeffery, S., Patton, S., and O, M. 2017. Application of a framework to assess vulnerability of biological components to ship-source oil spills in the marine environment in the Pacific Region. DFO Can. Sci. Advis. Sec. Res. Doc. 2017/057. ix + 145 p.
- Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning. *Elements* **1**: 337-387. doi:10.1007/b94608.
- Hawkins, B.A. 2012. Eight (and a half) deadly sins of spatial analysis. *J. Biogeogr.* **39**(1): 1-9.
- Hawkins, D.M., Basak, S.C., and Mills, D. 2003. Assessing model fit by cross-validation. *Journal of Chemical Information and Computer Sciences* **43**: 579-586. doi:10.1021/ci025626i.
- Hepinstall, J., Krohn, W., and Sader, S. 2002. Effects of niche width on the performance and agreement of avian habitat models. In *Predicting species occurrences. Issues of scale and accuracy.* Edited by J. Scott and P. Heglund and M. Morisson and J. Haufler and M. Raphael and W. Wall and F. Samson. Island Press, Washington. pp. 593-606.
- Hernandez, P.A., Graham, C.H., Master, L.L., and Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**: 773-785.
- Hijmans, R.J. 2012. Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. *Ecology* **93**: 679-688. doi:10.1890/11-0826.1.
- Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J. 2017. *dismo: Species Distribution Modeling.* R package version 1.1-4.
- Hirzel, A.H., and Le Lay, G. 2008. Habitat suitability modelling and niche theory. *J. Appl. Ecol.* **45**(5): 1372-1381.
- Hirzel, A.H., Hausser, J., Chessel, D., and Perrin, N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* **83**(7): 2027-2036.
- Hirzel, A.H., Lay, G.L., Helfer, V., Randin, C., and Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. *Ecol. Model.* **199**: 142-152.
- Hothorn, T., Müller, J., Schröder, B., Kneib, T., and Brandl, R. 2011. Decomposing environmental, spatial, and spatiotemporal components of species distributions. *Ecol. Monogr.* **81**: 329-347. doi:10.1890/10-0602.1.
- Hu, C., Lee, Z., and Franz, B. 2012. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *J. Geophys. Res. (C Oceans)* **117**: 1-25. doi:10.1029/2011JC007395.
- Hughes, J., Haran, M., and Caragea, P.C. 2011. Autologistic models for binary data on a lattice. *Environmetrics* **22**: 857-871. doi:10.1002/env.1102.
- Iturbide, M., Bedia, J., and Gutiérrez, J. 2018. Tackling Uncertainties of Species Distribution Model Projections with Package *mopa*. *R Journal* **10**(1).
- Jones-Farrand, D.T., Fearer, T.M., Thogmartin, W.E., Thompson, F.R.I., Nelson, M.D., and Tirpak, J.M. 2011. Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble prediction. *Ecol. Appl.* **21**: 2269-2282. doi:10.1890/10-1047.1.
- Jones, C.C., Acker, S.A., and Halpern, C.B. 2010. Combining local-and large-scale models to predict the distributions of invasive plant species. *Ecol. Appl.* **20**(2): 311-326.

-
- Jones, M.C., and Cheung, W.W.L. 2015. Multi-model ensemble projections of climate change effects on. *ICES J. Mar. Sci.* **72**: 741-752.
- Kaschner, K., Watson, R., Trites, A.W., and Pauly, D. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Mar. Ecol. Prog. Ser.* **316**: 285-310.
- Keitt, T.H., Bjørnstad, O.N., Dixon, P.M., and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. *Ecography* **25**(5): 616-625.
- Kühn, I., and Dormann, C.F. 2012. Less than eight (and a half) misconceptions of spatial analysis. *J. Biogeogr.* **39**(5): 995-998.
- Landis, J.R., and Koch, G.G. 1977. The measurement of observer agreement for categorical data. *Biometrics*: 159-174.
- Lawson, C.R., Hodgson, J.A., Wilson, R.J., and Richards, S.A. 2014. Prevalence, thresholds and the performance of presence-absence models. *Methods in Ecology and Evolution* **5**(1): 54-64.
- Leathwick, J., Elith, J., Francis, M., Hastie, T., and Taylor, P. 2006. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Mar. Ecol. Prog. Ser.* **321**: 267-281.
- Leathwick, J., Elith, J., Chadderton, W., Rowe, D., and Hastie, T. 2008. Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *J. Biogeogr.* **35**(8): 1481-1497.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* **74**(6): 1659-1673.
- Legendre, P., and Fortin, M. 1989. Spatial pattern and ecological analysis. *Plant Ecol.* **80**: 107-138.
- Lessard, J., and Campbell, A. 2007. Describing northern abalone, *Haliotis kamtschatkana*, habitat: focusing rebuilding efforts in British Columbia, Canada. *J. Shellfish Res.* **26**(3): 677-686.
- Levin, S.A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology* **73**(6): 1943-1967.
- Levins, R. 1966. The strategy of model building in population biology. *Am. Sci.* **54**(4): 421-431.
- Li, J., Heap, A.D., Potter, A., and Daniell, J.J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environ. Model. Software* **26**(12): 1647-1659.
- Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. *R news* **2**(3): 18-22.
- Lichstein, J.W., Simons, T.R., Shiner, S.A., Franzreb, K.E., Monographs, S.E., and Aug, N. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecol. Monogr.* **72**: 445-463. doi:10.1890/0012-9615(2002)072[0445:saaami]2.0.co;2.
- Link, J.S., Ihde, T.F., Harvey, C.J., Gaichas, S.K., Field, J.C., Brodziak, J.K.T., Townsend, H.M., and Peterman, R.M. 2012. Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. *Prog. Oceanogr.* **102**: 102-114.
- Link, W.A., and Barker, R.J. 2006. Ecological Society of America Model Weights and the Foundations of Multimodel Inference. *Ecology* **87**: 2626-2635.
-

-
- Lobo, J.M., Jiménez-Valverde, A., and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology & Biogeography* **17**: 145-151.
- Lobo, J.M., Jiménez-Valverde, A., and Hortal, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. *Ecography* **33**(1): 103-114.
- McPherson, J.M., and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography* **30**(1): 135-151.
- MacArthur, R. 1955. Fluctuations of animal populations and a measure of community stability. *Ecology* **36**(3): 533-536.
- Manel, S., Williams, H.C., and Ormerod, S.J. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* **38**: 921-931.
- Martin, M.P., Orton, T.G., Lacarce, E., Meersmans, J., Saby, N.P.A., Paroissien, J.B., Jolivet, C., Boulonne, L., and Arrouays, D. 2014. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. *Geoderma* **223-225**: 97-107. doi:10.1016/j.geoderma.2014.01.005.
- Masson, D., and Fine, I. 2012. Modeling seasonal to interannual ocean variability of coastal British Columbia. *J. Geophys. Res. (C Oceans)* **117**(C10): C10019:10011-10014. doi:10.1029/2012jc008151.
- Maxwell, S.M., Hazen, E.L., Lewison, R.L., Dunn, D.C., Bailey, H., Bograd, S.J., Briscoe, D.K., Fossette, S., Hobday, A.J., and Bennett, M. 2015. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. *Mar. Policy* **58**: 42-50.
- Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., and Vanaverbeke, J. 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecol. Model.* **222**(3): 588-597.
- Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., and Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? *Ecography* **37**(12): 1267-1281.
- MPO. 2009. Élaboration d'un cadre et de principes pour la classification biogéographique des zones marines canadiennes. *Secr. can. de consult. sci. du MPO, Avis sci.* 2009/056.
- Oppel, S., Gardner, B., O'Connell, A.F., Louzao, M., Miller, P.I., Meirinho, A., and Ramírez, I. 2011. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biol. Conserv.* **156**: 94-104. doi:10.1016/j.biocon.2011.11.013.
- Ord, J.K. 1979. Time-series and spatial patterns in ecology. In *Spatial and Temporal Analysis in Ecology*. Edited by R.M. Cormack and J.K. Ord. International Cooperative Publishing House, Fairland, Maryland, USA.
- Pearce, J., and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. *Ecol. Model.* **133**: 225-245.
- Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., and Lees, D.C. 2006. Model-based uncertainty in species range prediction. *J. Biogeogr.* **33**(10): 1704-1711.
- Phillips, S.J., and Elith, J. 2010. POC plots: calibrating species distribution models with presence-only data. *Ecology* **91**(8): 2476-2484.

-
- Planque, B., Loots, C., Petitgas, P., Lindstrøm, U., and Vaz, S. 2011. Understanding what controls the spatial distribution of fish populations using a multi-model approach. *Fish. Oceanogr.* **20**(1): 1-17.
- R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Raleigh, R.F., Miller, W.J., and Nelson, P.C. 1986. Habitat suitability index models and instream flow suitability curves: Chinook Salmon. U.S. Fish And Wildlife Service. Biological Report **82**(10.122).
- Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., and Guisan, A. 2006. Are niche-based species distribution models transferable in space? *J. Biogeogr.* **33**(10): 1689-1703.
- Ridgeway, G. 2007. Generalized Boosted Models: A Guide to the gbm Package. In R package vignette.
- Roberts, D.R., Bahn, V., Ciuti, S., Wintle, B.A., Guillera-Aroita, G., Elith, J., Warton, D.I., Hartig, F., Dormann, C.F., Lahoz-Monfort, J.J., Hauenstein, S., Thuiller, W., Schröder, B., and Boyce, M.S. 2016. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**: 913-929. doi:10.1111/ecog.02881.
- Robinson, N.M., Lundquist, C.J., Sutherland, J.E., Nelson, W.A., and Costello, M.J. 2017. A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice. *Frontiers in Marine Science* **4**: 1-11. doi:10.3389/fmars.2017.00421.
- Rooper, C.N., Zimmermann, M., and Prescott, M.M. 2017. Comparison of modeling methods to predict the spatial distribution of deep-sea coral and sponge in the Gulf of Alaska. *Deep Sea Res. (I Oceanogr. Res. Pap.)* **126**: 148-161.
- Santoul, F., Mengin, N., Cereghino, R., Figuerola, J., and Mastrorillo, S. 2005. Environmental factors influencing the regional distribution and local density of a small benthic fish: the stone loach (*Barbatula barbatula*). *Hydrobiologia* **544**(1): 347-355.
- Sappington, J.M., Longshore, K.M., and Thompson, D.B. 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. *The Journal of wildlife management* **71**(5): 1419-1426.
- Schwarz, G. 1978. Estimating the dimension of a model. *The annals of statistics* **6**(2): 461-464.
- Segurado, P., and Araújo, M.B. 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* **31**: 1555-1568.
- Silvert, W.L. 1981. Principles of ecosystem modelling. In *Analysis of Marine Ecosystems*. Edited by A.R. Longhurst. Academic Press, New York. pp. 651-676.
- Soberón, J., and Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* **2**: 1-10.
- Soberón, J., and Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. *PNAS* **106**(Supplement 2): 19644-19650.
- Stockwell, D.R., and Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* **148**(1): 1-13.
- Tabachnick, B.G., Fidell, L.S., and Ullman, J.B. 2007. *Using multivariate statistics*. Pearson, Boston, MA.
-

-
- Tien, N., Craeymeersch, J., Van Damme, C., Couperus, A., Adema, J., and Tulp, I. 2017. Burrow distribution of three sandeel species relates to beam trawl fishing, sediment composition and water velocity, in Dutch coastal waters. *J. Sea Res.* **127**: 194-202.
- Trachsel, M., and Telford, R.J. 2016. Technical note: Estimating unbiased transfer-function performances in spatially structured environments. *Climate of the Past* **12**: 1215-1223. doi:10.5194/cp-12-1215-2016.
- Tuanmu, M.N., Viña, A., Roloff, G.J., Liu, W., Ouyang, Z., Zhang, H., and Liu, J. 2011. Temporal transferability of wildlife habitat models: implications for habitat monitoring. *J. Biogeogr.* **38**(8): 1510-1523.
- USFWS. 1981. Standards for the development of habitat suitability index models. U.S. Department of Interior, Fish and Wildlife Service, Division of Ecological Services ESM 103. 162 + viii pp.
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., and Guillera-Aroita, G. 2018. blockVC: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*: 1-8. doi:10.1111/2041-210X.13107.
- Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. 2018. Unified geomorphological analysis workflows with benthic terrain modeler. *Geosciences* **8**(3): 94.
- Wells, R.D., Boswell, K.M., Cowan Jr., J.H., and Patterson III, W.F. 2008. Size selectivity of sampling gears targeting red snapper in the northern Gulf of Mexico. *Fisheries Research* **89**(3): 294-299.
- Wenger, S.J., and Olden, J.D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution* **3**(2): 260-267.
- Westerbom, M., and Jattu, S. 2006. Effects of wave exposure on the sublittoral distribution of blue mussels *Mytilus edulis* in a heterogeneous archipelago. *Mar. Ecol. Prog. Ser.* **306**: 191-200.
- Wiens, J.A. 1989. Spatial scaling in ecology. *Funct. Ecol.* **3**: 385-397.
- Wilson, E.O., and MacArthur, R.H. 1967. *The theory of island biogeography*. Princeton, NJ.
- Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., and Barbosa, A.M. 2018. Outstanding challenges in the transferability of ecological models. *Trends Ecol. Evol.* **33**(10):790-802.

ANNEXE A. PRÉDICTIONS DE LA RÉPARTITION DE DOUZE ESPÈCES À PARTIR DE L'APPLICATION DU CADRE

Cette annexe présente les résultats de l'application du cadre décrit à la section 4 à douze espèces. Pour chaque espèce, nous présentons des figures illustrant :

1. la distribution des observations de présence et d'absence utilisées pour la modélisation;
2. les prédictions de la répartition de l'espèce et leur incertitude;
3. l'influence relative et les effets marginaux des variables prédictives environnementales.

L'incertitude est présentée dans l'espace en montrant, d'une part, la différence entre la prédiction du modèle d'indice de qualité de l'habitat (IQH) fondé sur les connaissances et la prédiction du modèle d'ensemble, et de l'autre, la variation entre plusieurs prédictions d'ensemble au moyen de l'écart-type (voir la section 3.6). Chaque modèle d'ensemble combinait les prédictions d'un modèle linéaire généralisé (MLG) et d'un modèle d'arbre de régression stimulée (ARS).

Pour les espèces sur lesquelles on ne disposait pas de données appropriées pour construire des modèles corrélatifs, l'étoile ocrée (*Pisaster ochraceus*), le complexe de moules bleues (*Mytilus edulis*, *M. trossulus* et *M. galloprovincialis*), la palourde du Pacifique (*Leukoma staminea*) et la pennatule orangée (*Ptilosarcus gurneyi*), nous présentons un seul modèle IQH sans mesure de l'incertitude.

Nous ne pouvons pas afficher les observations de présence ou les répartitions prédites pour l'ormeau nordique en raison de son statut d'espèce en péril (LEP). Des restrictions concernant la communication de l'information sur l'habitat connu de l'ormeau sont en place pour protéger les populations de l'espèce en voie de rétablissement. Voir les relations entre l'influence relative et les effets marginaux à la Figure 4.3.

Panope du Pacifique

Panopea generosa

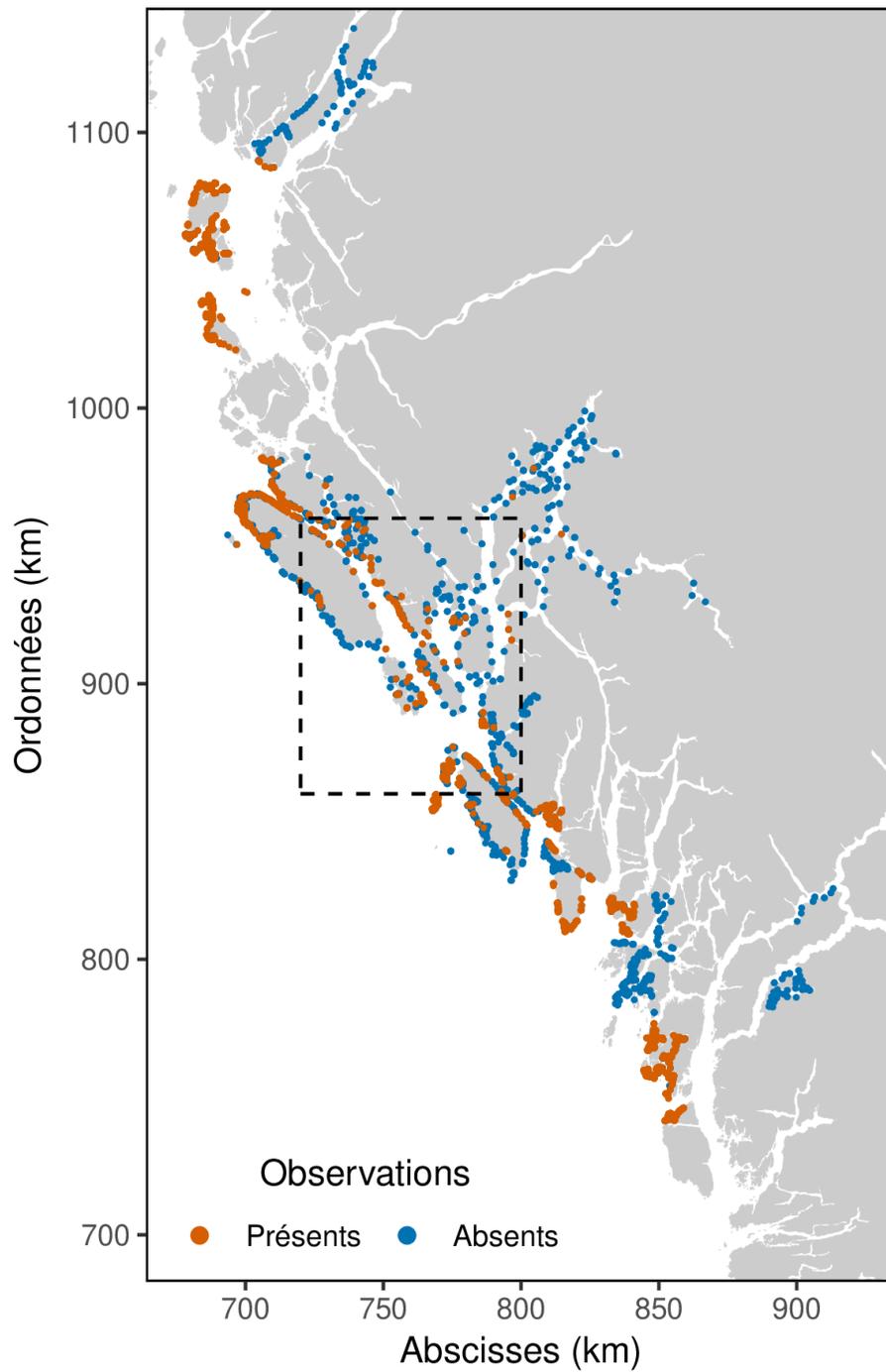


Figure A.1. Observations de la présence et de l'absence de la panope dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.2.

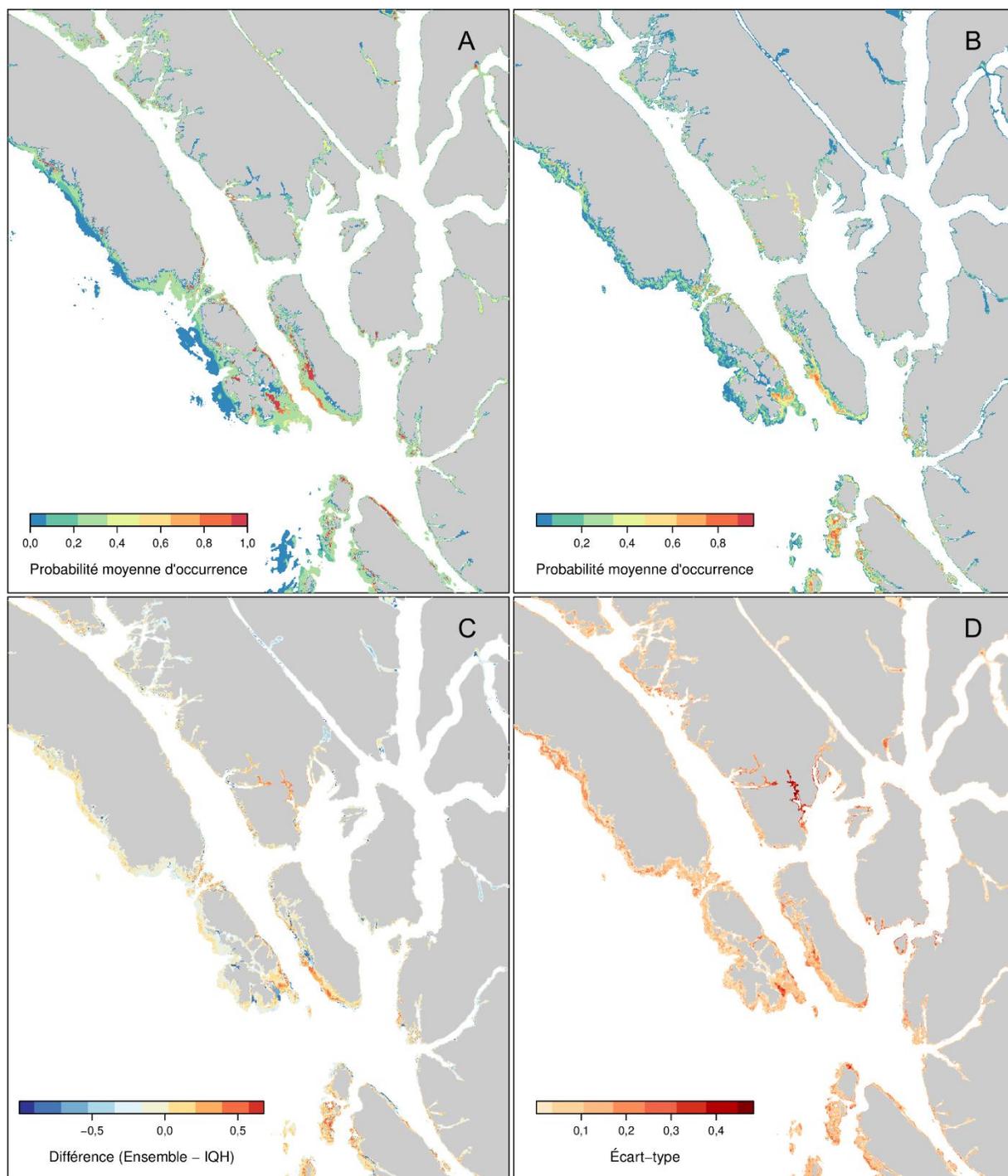
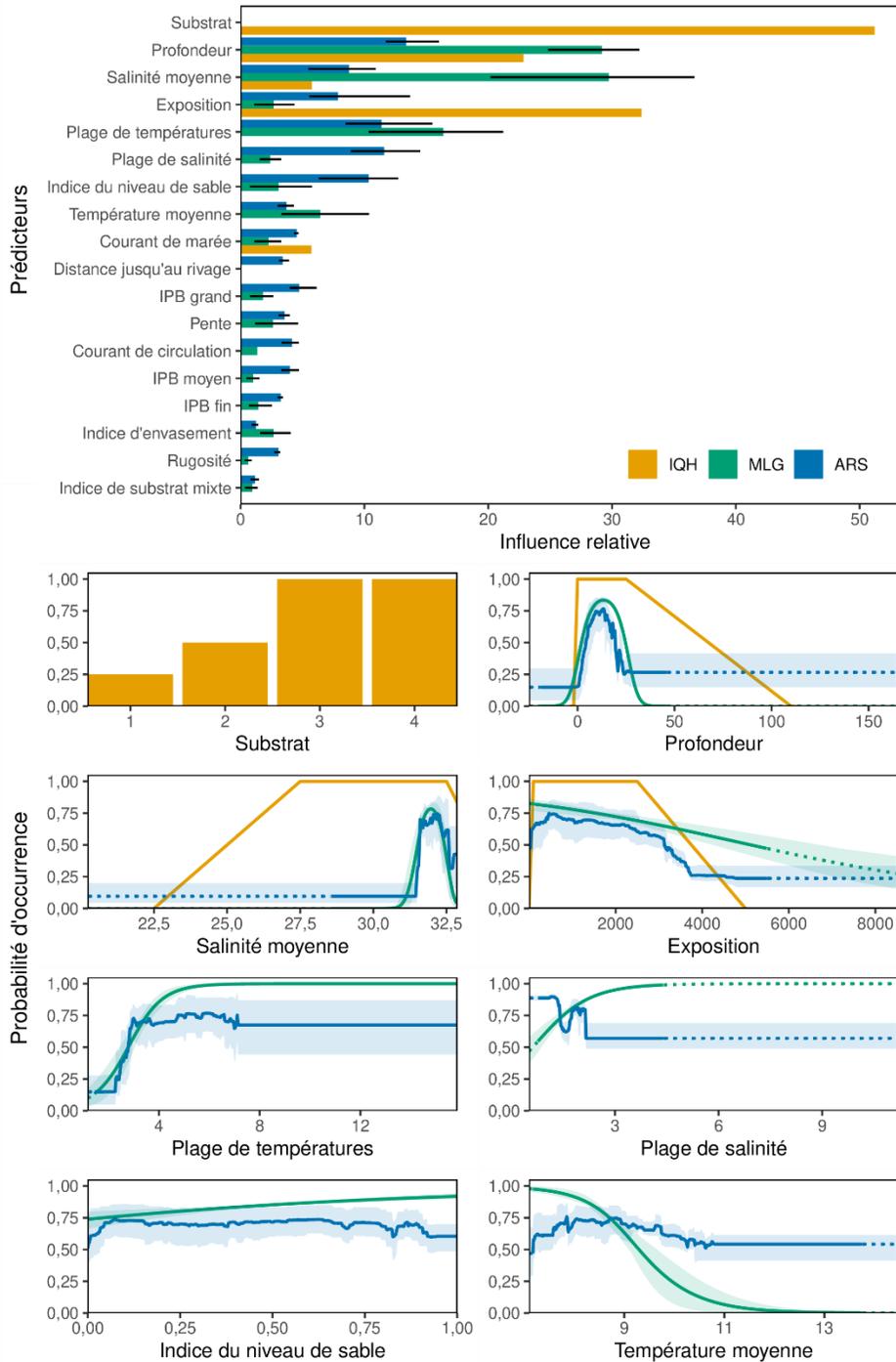


Figure A.2. Prédications de la répartition de la panope et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.



Oursin rouge

Mesocentrotus franciscanus

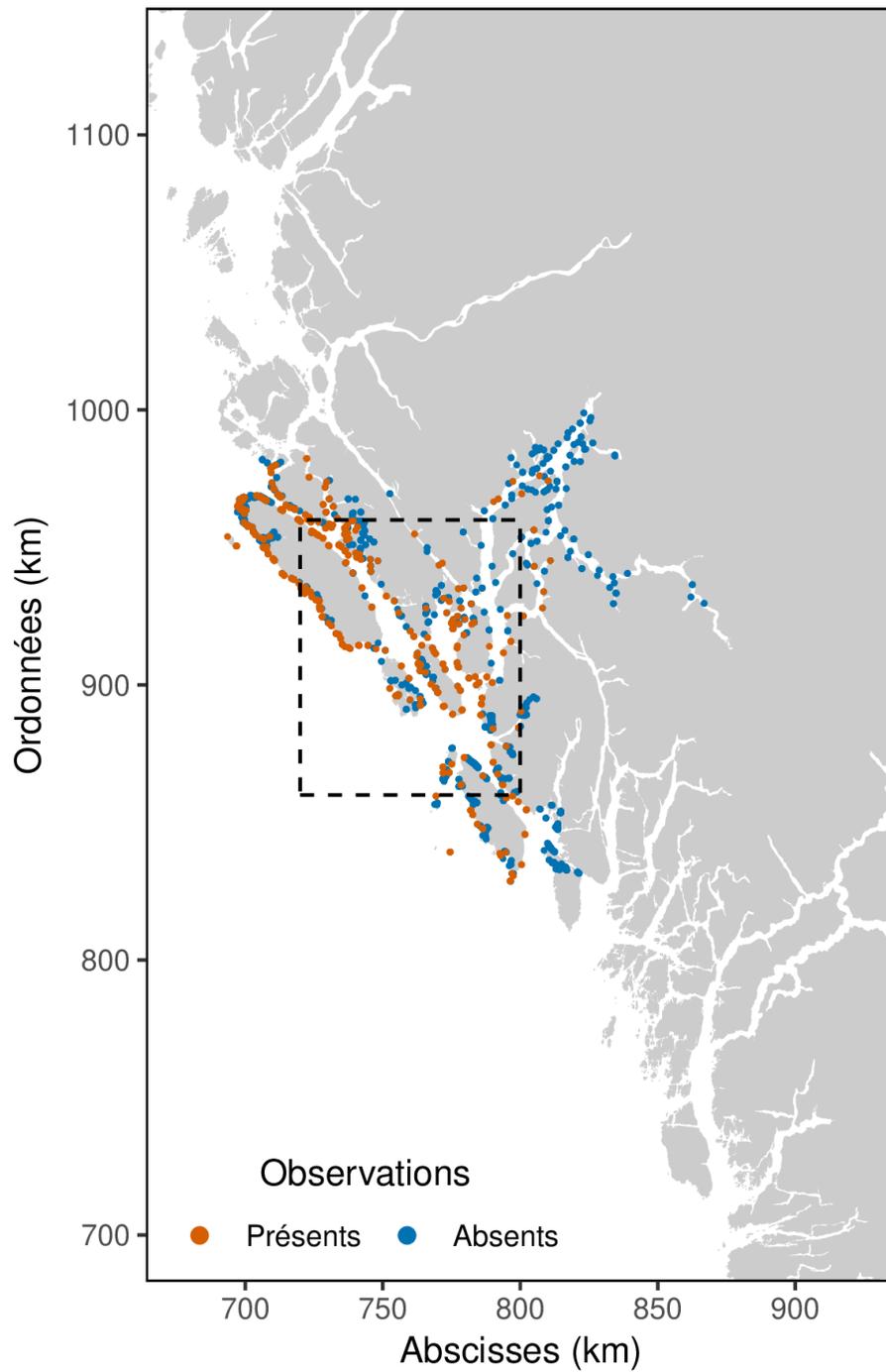


Figure A.4. Observations de la présence et de l'absence de l'oursin rouge dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.5.

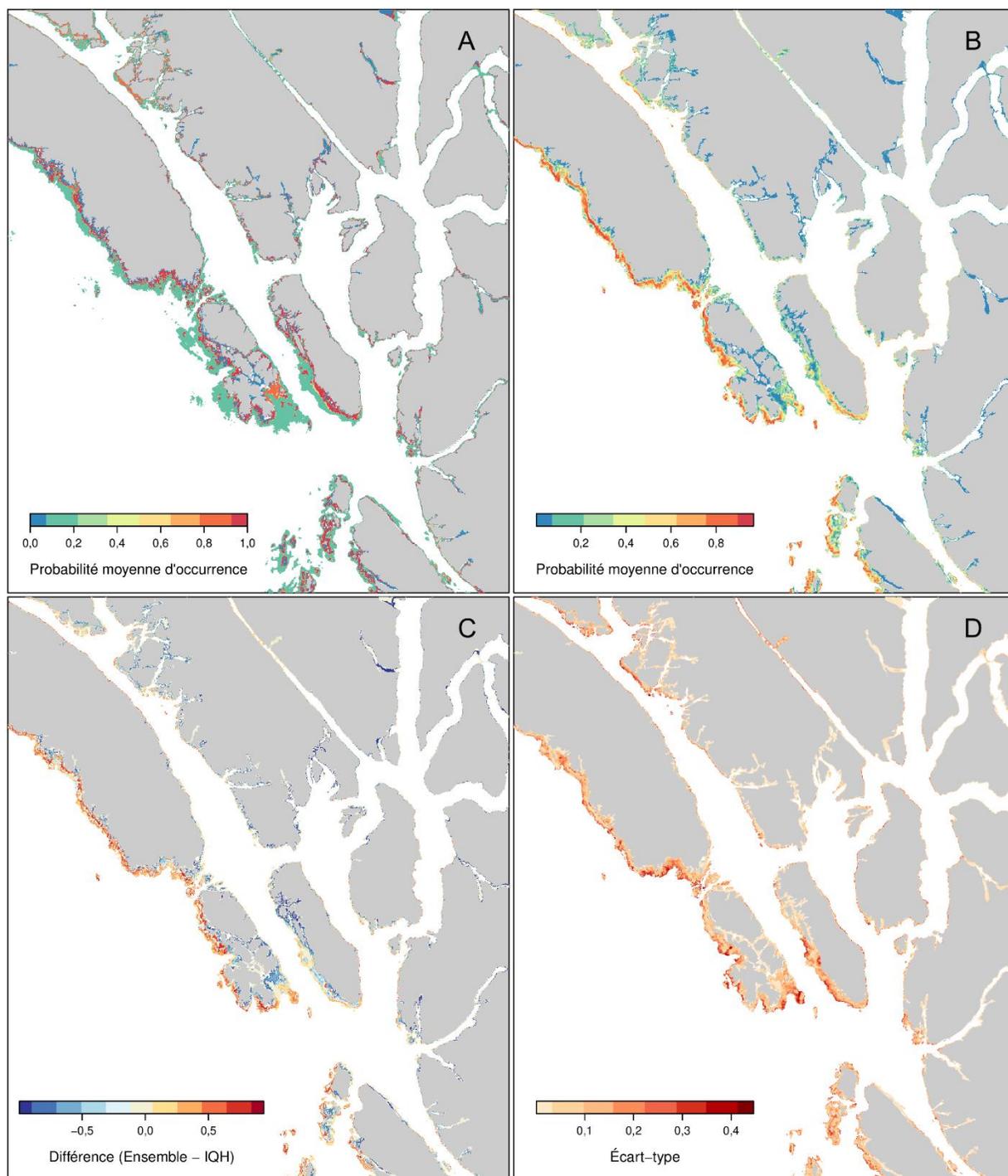


Figure A.5. Prédications de la répartition de l'oursin rouge et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée à la figure A.4 par la case en pointillés.

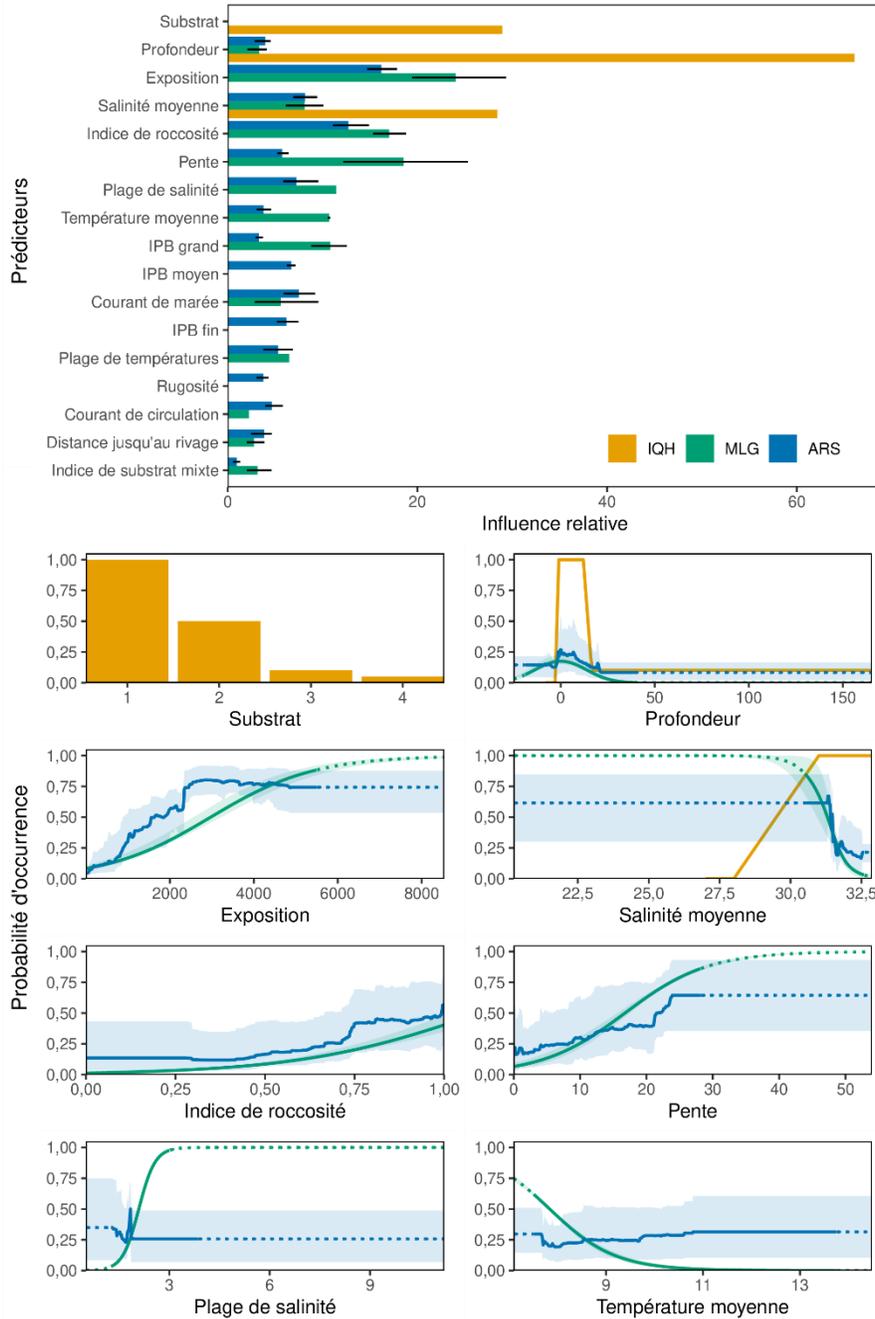


Figure A.6. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS de l'oursin rouge. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.

Varech pterygophora

Pterygophora californica

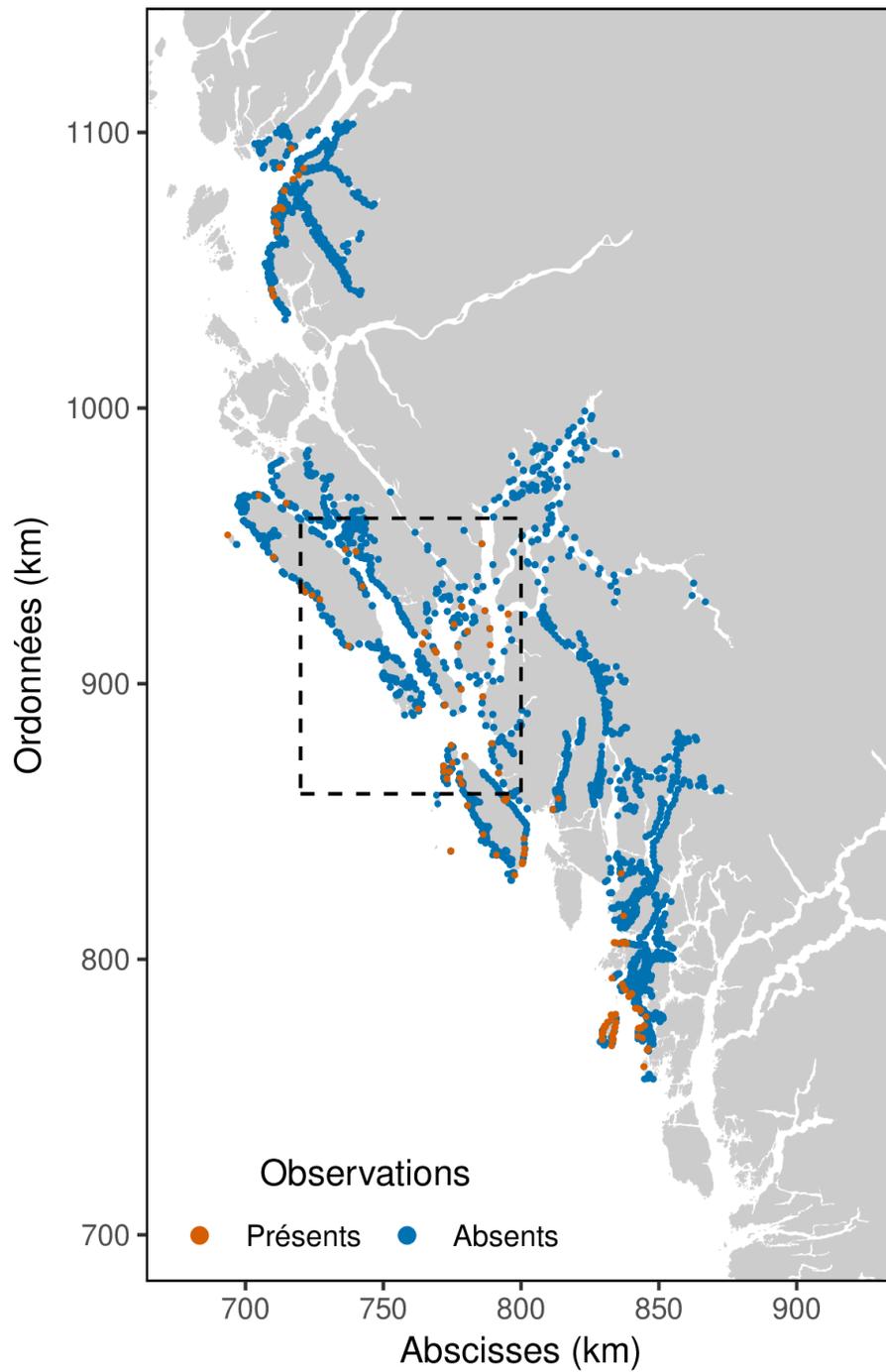


Figure A.7. Observations de la présence et de l'absence du varech *Pterygophora* dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée sur la figure A.8.

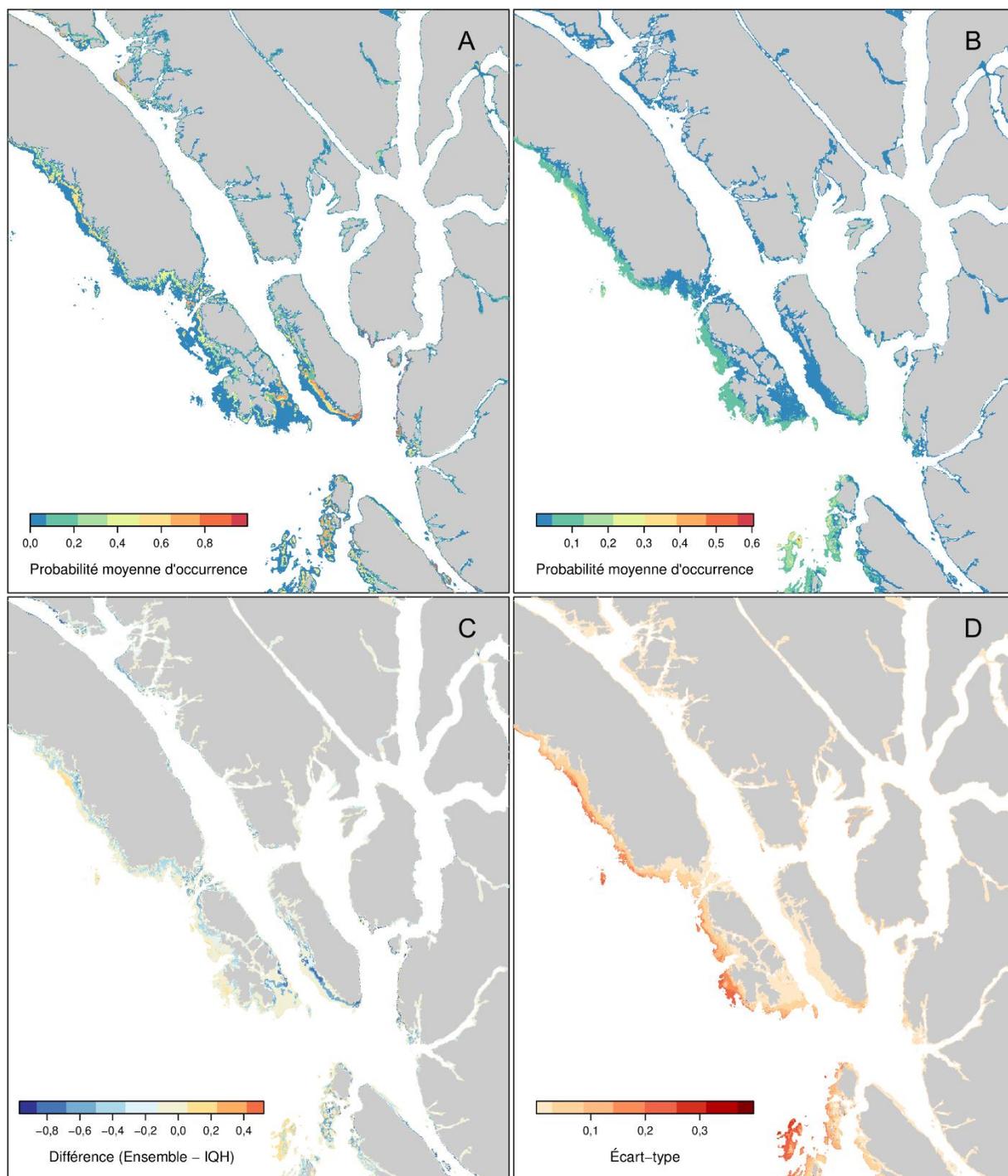


Figure A.8. Prédications de la répartition du varech *Pterygophora* et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée sur la figure A.7 par la case en pointillés.

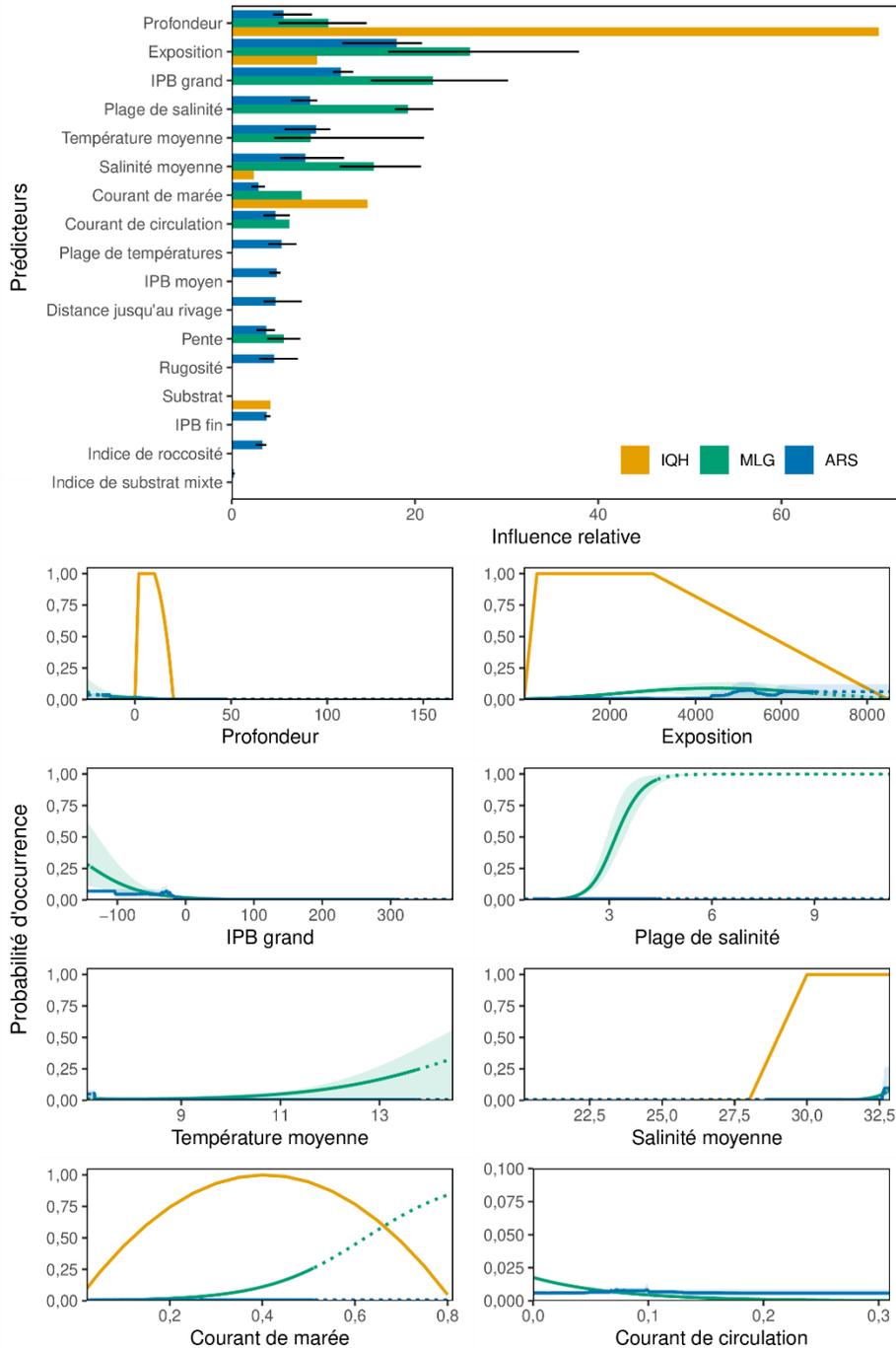


Figure A.9. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du varech *Pterygophora*. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.

Zostère marine

Zostera spp.

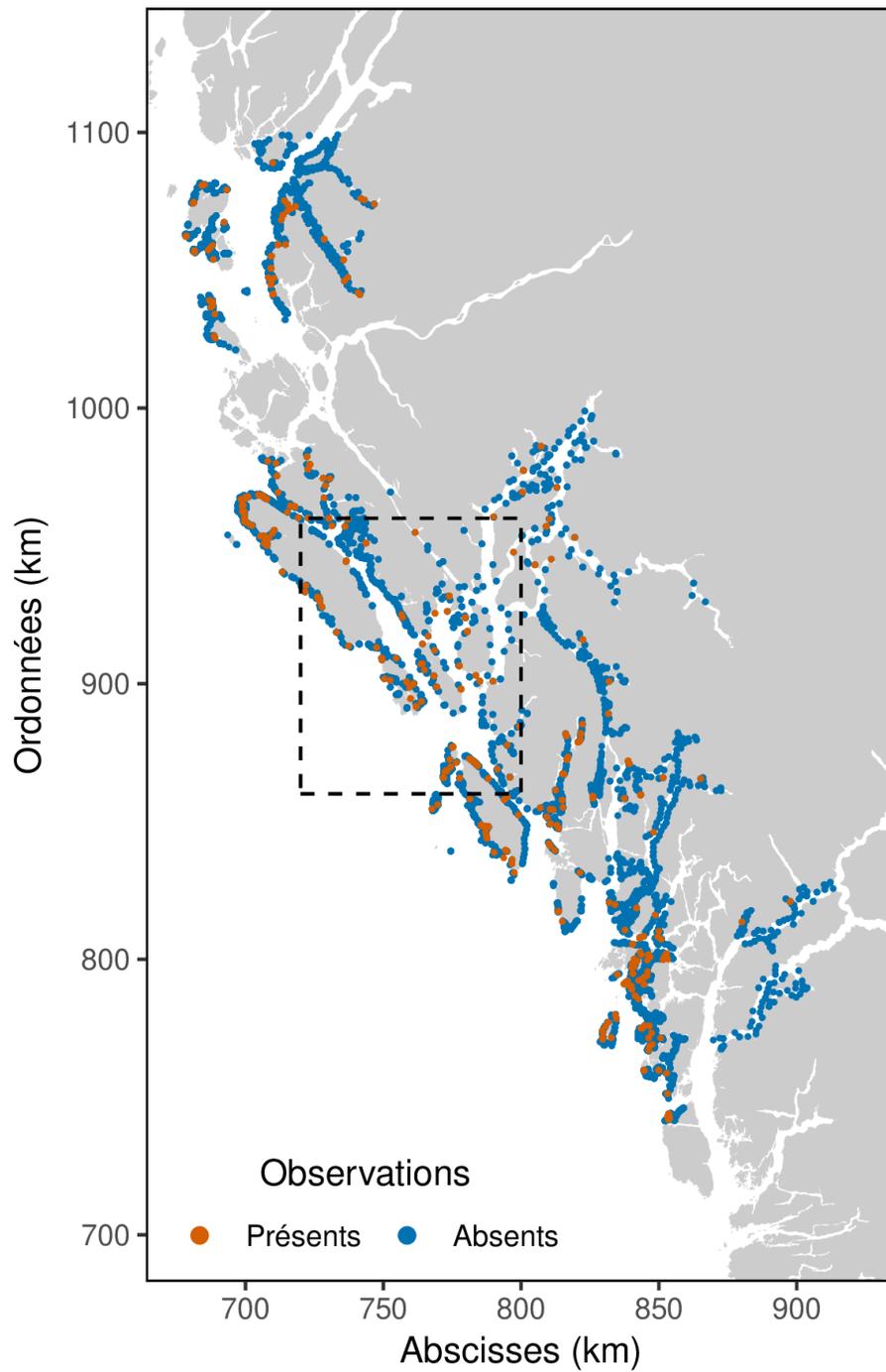


Figure A.10. Observations de la présence et de l'absence de la zostère marine dans la zone d'étude des eaux côtières. La case en pointillés indique la zone représentée à la figure A.11.

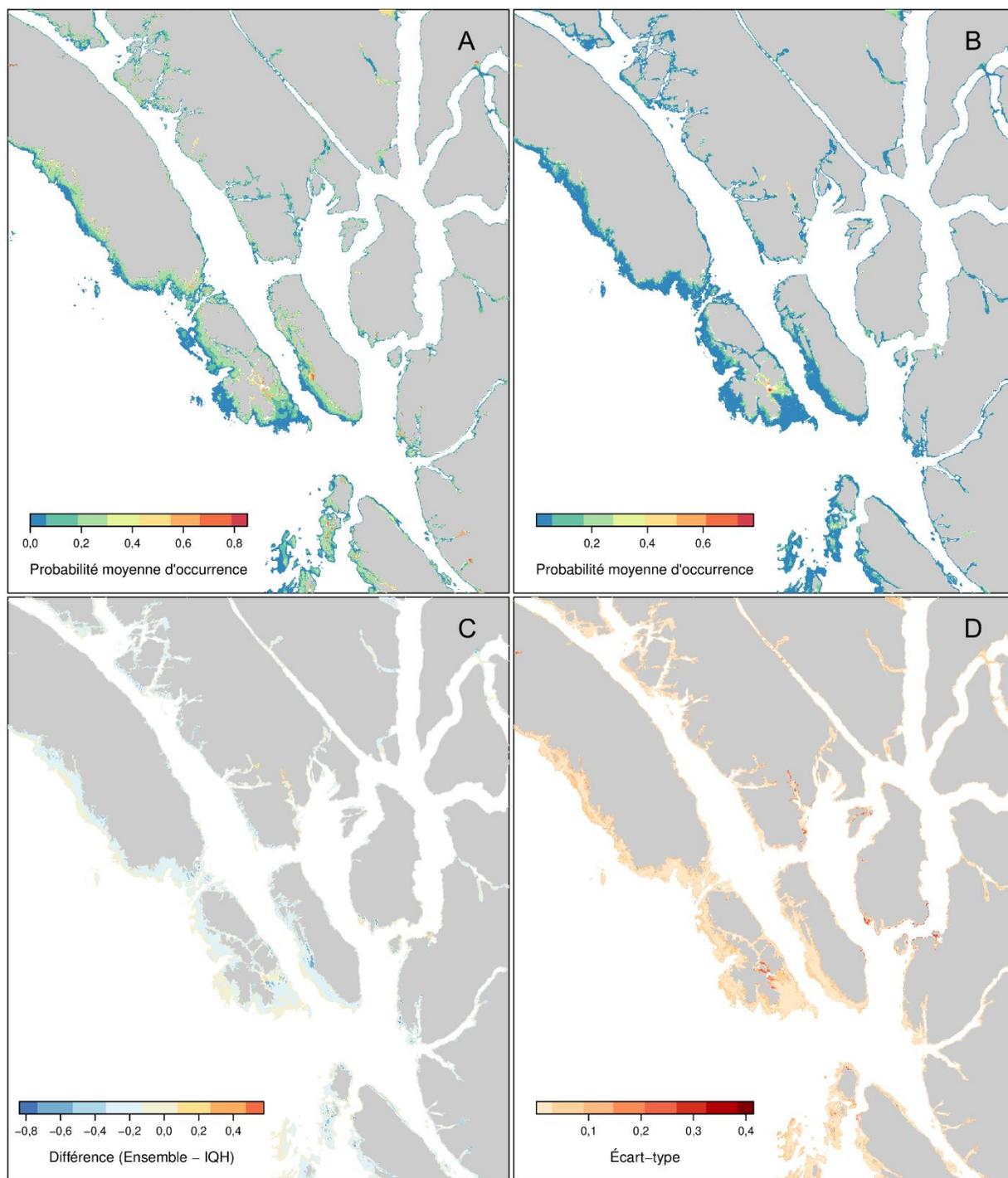


Figure A.11. Prédiction de la répartition de la zostère marine et incertitude correspondante. Prédiction de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble. La zone représentée ici est indiquée sur la figure A.10 par la case en pointillés.

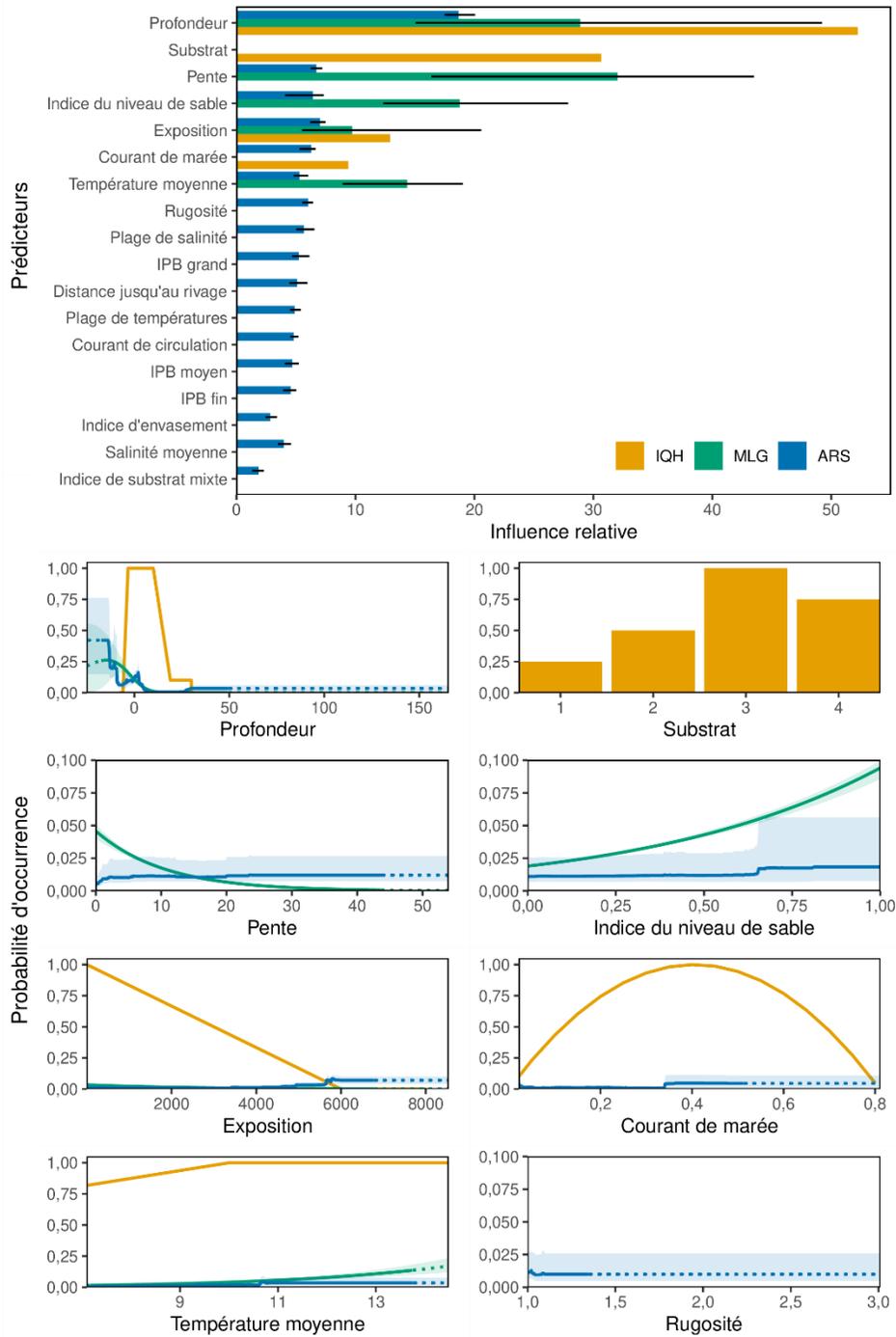


Figure A.12. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS de la zostère marine. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice continu du niveau de sable pour les modèles MLG et ARS.

Crabe dormeur

Metacarcinus magister

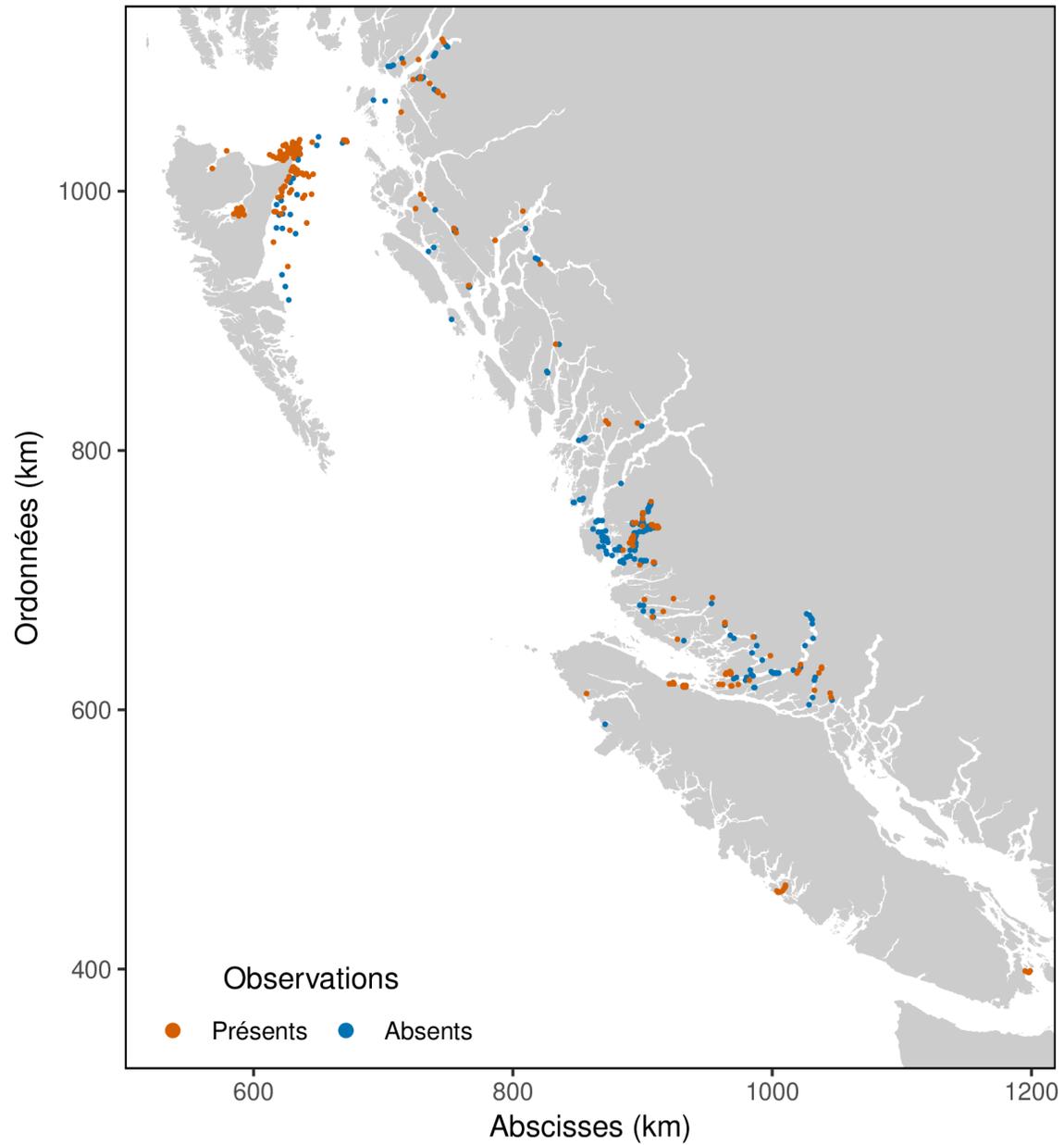


Figure A.13. Observations de la présence et de l'absence du crabe dormeur dans la zone d'étude du plateau.

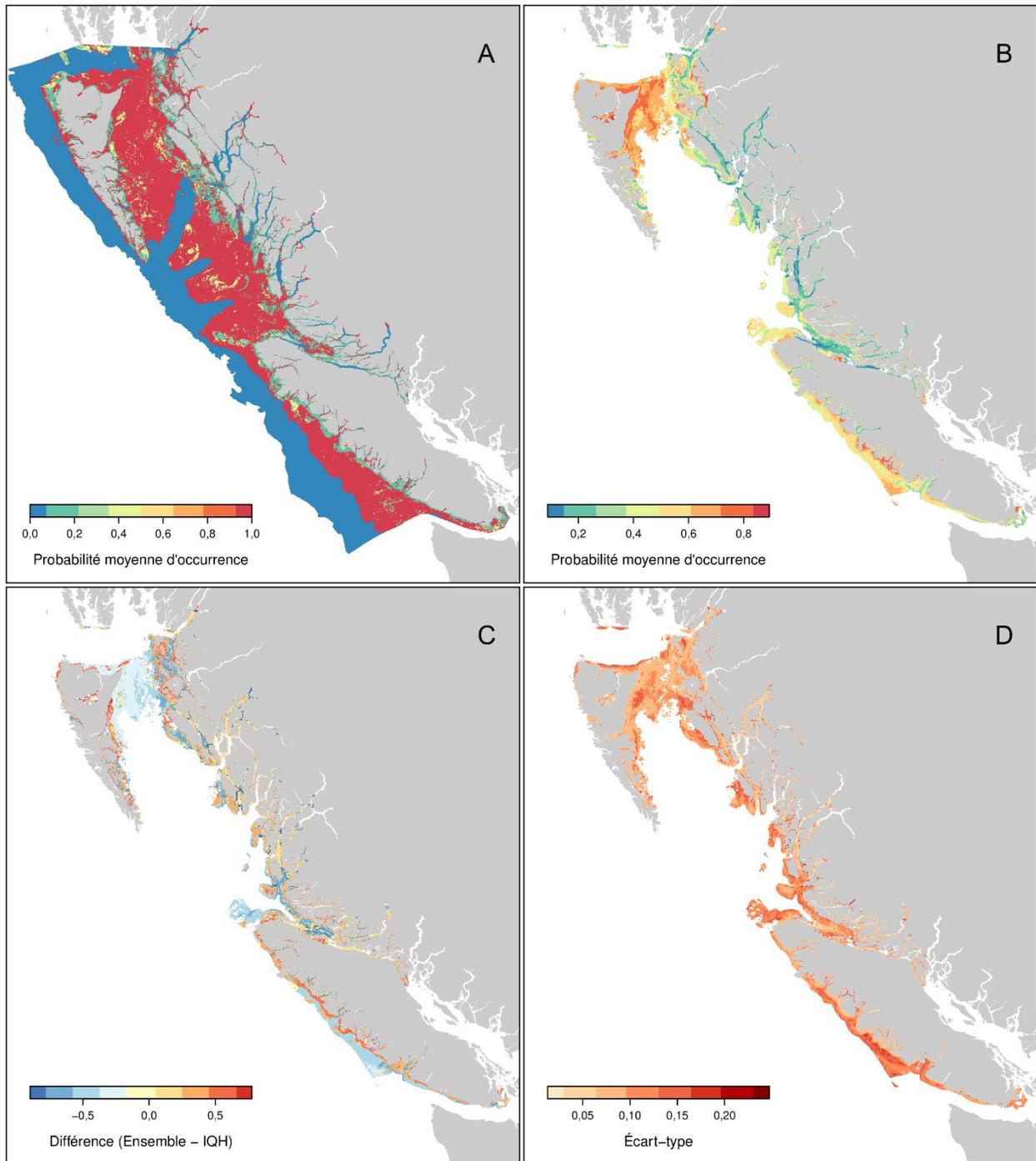


Figure A.14. Prédications de la répartition du crabe dormeur et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.

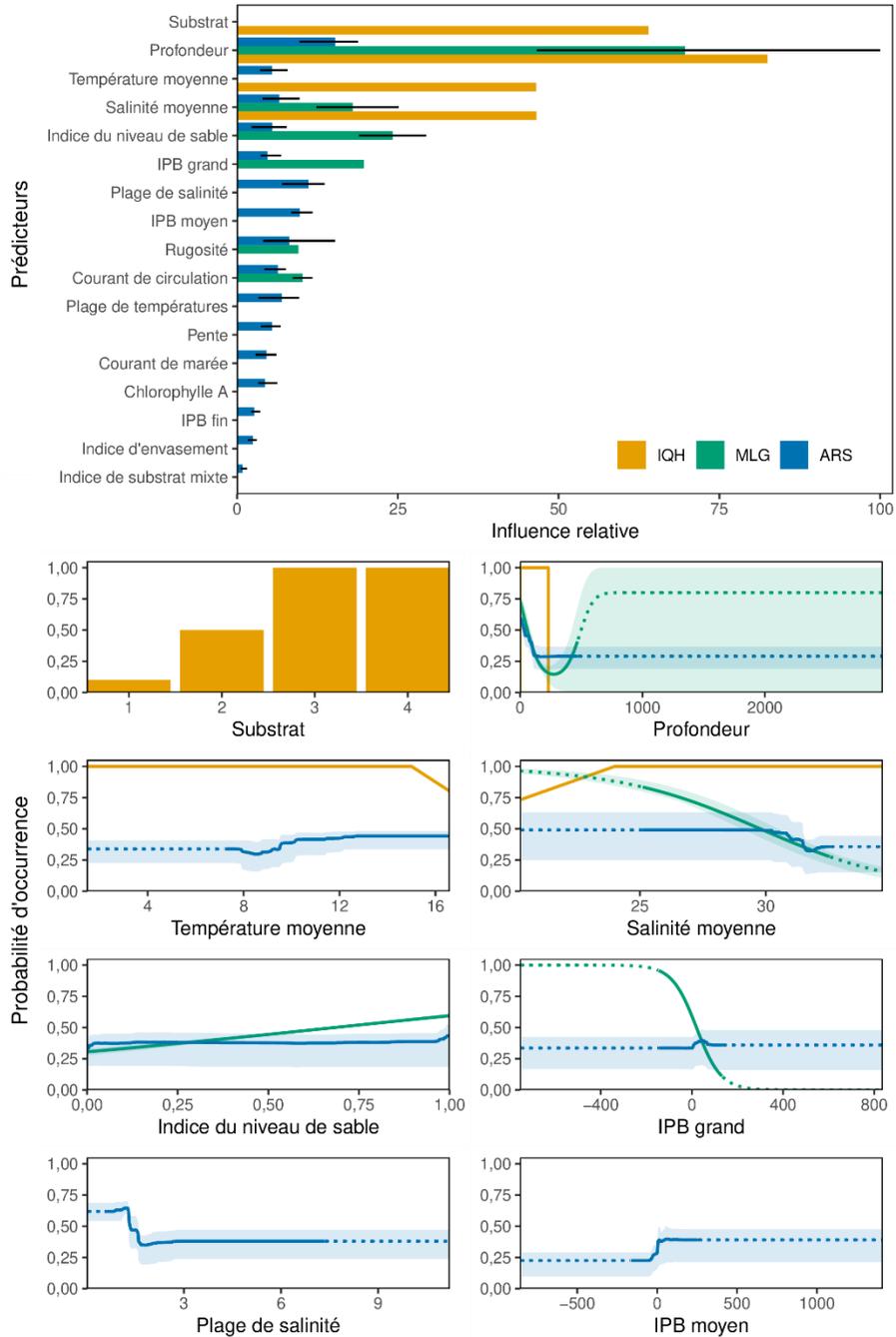


Figure A.15. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du crabe dormeur. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice continu du niveau de sable pour les modèles MLG et ARS.

Sébaste à dos épineux

Sebastes maliger

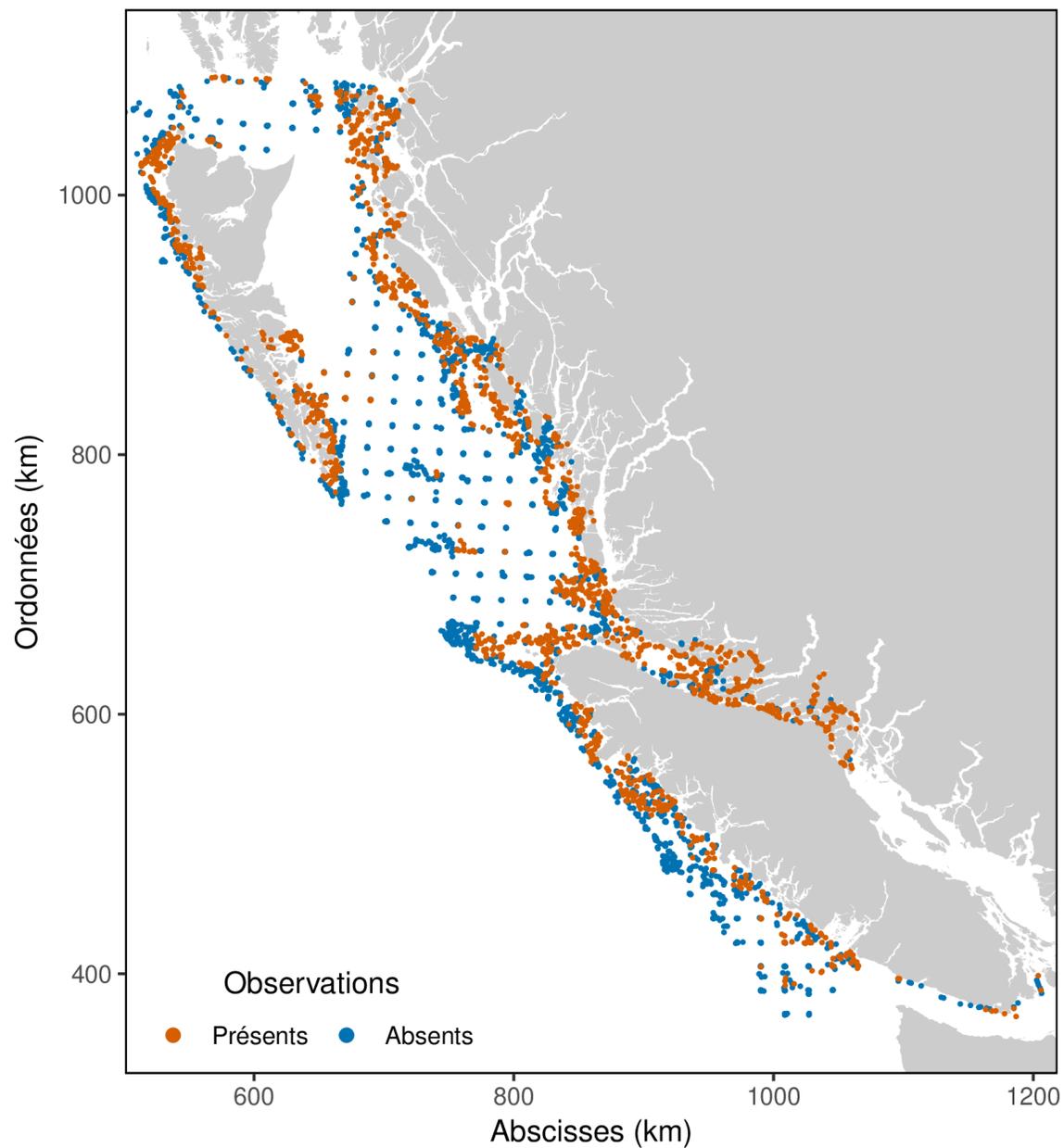


Figure A.16 Observations de la présence et de l'absence du sébaste à dos épineux dans la zone d'étude du plateau.

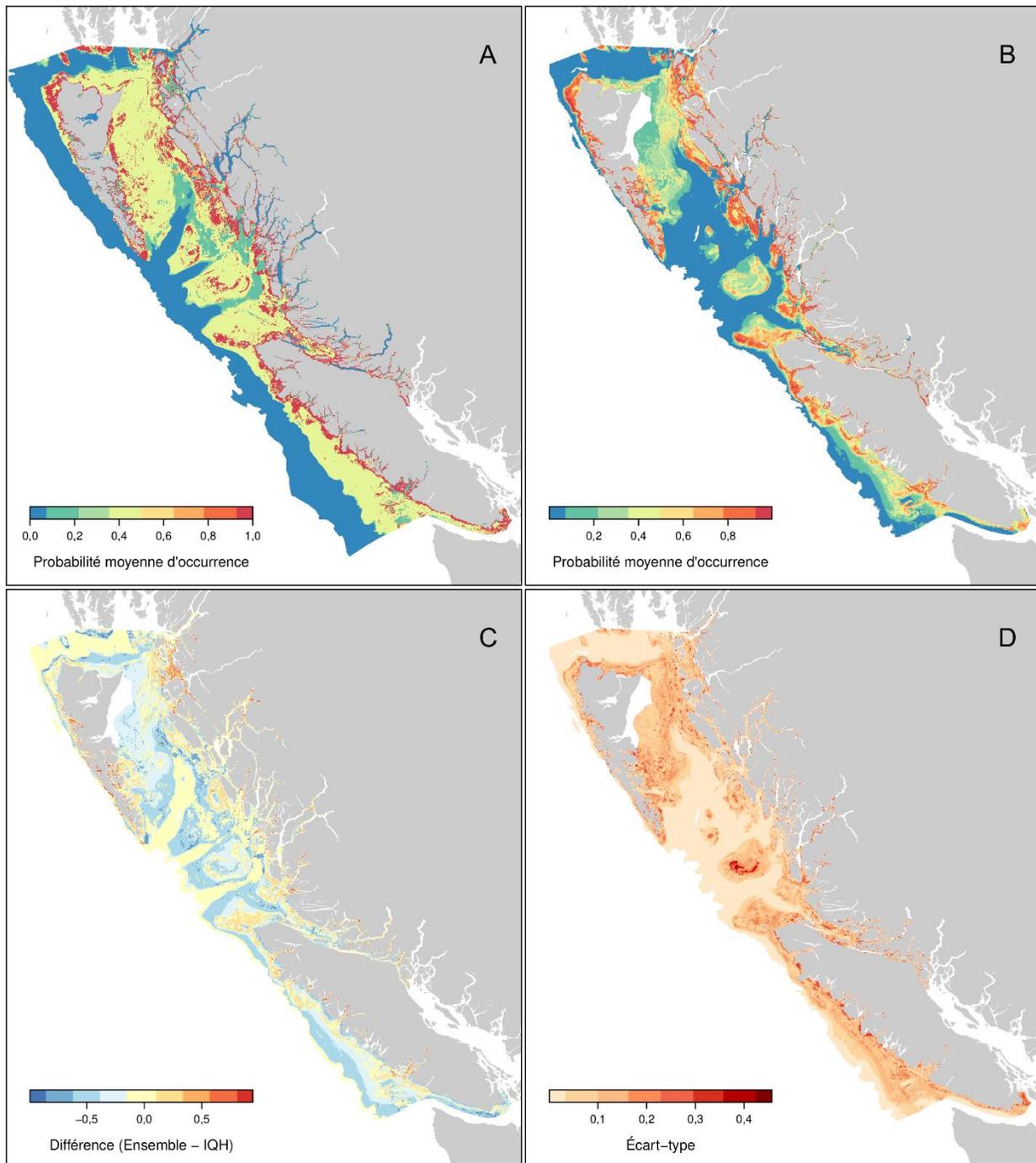


Figure A.17. Prédications de la répartition du sébaste à dos épineux et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.

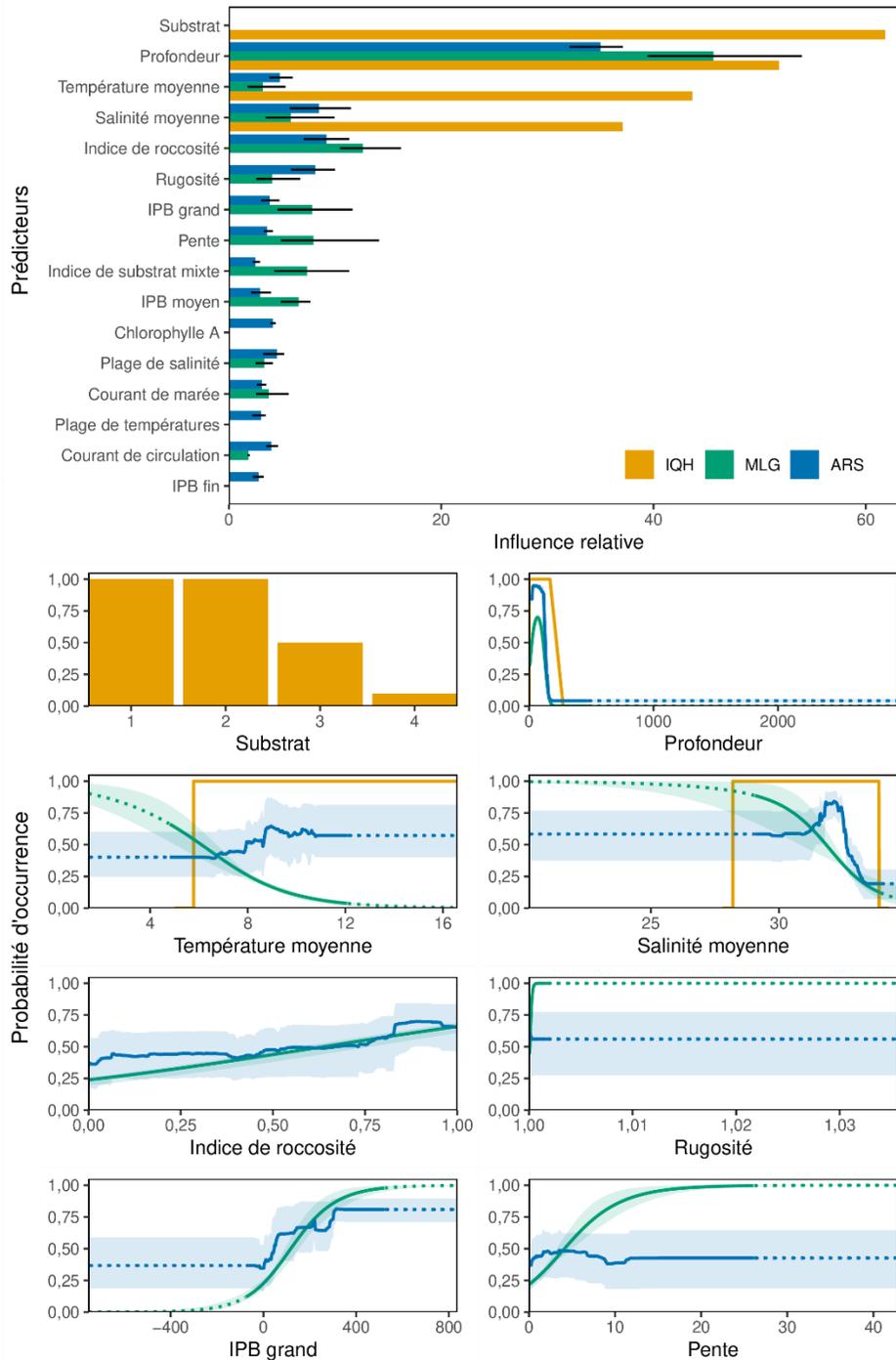


Figure A.18. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du sébaste à dos épineux. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.

Sébaste aux yeux jaunes

Sebastes ruberrimus

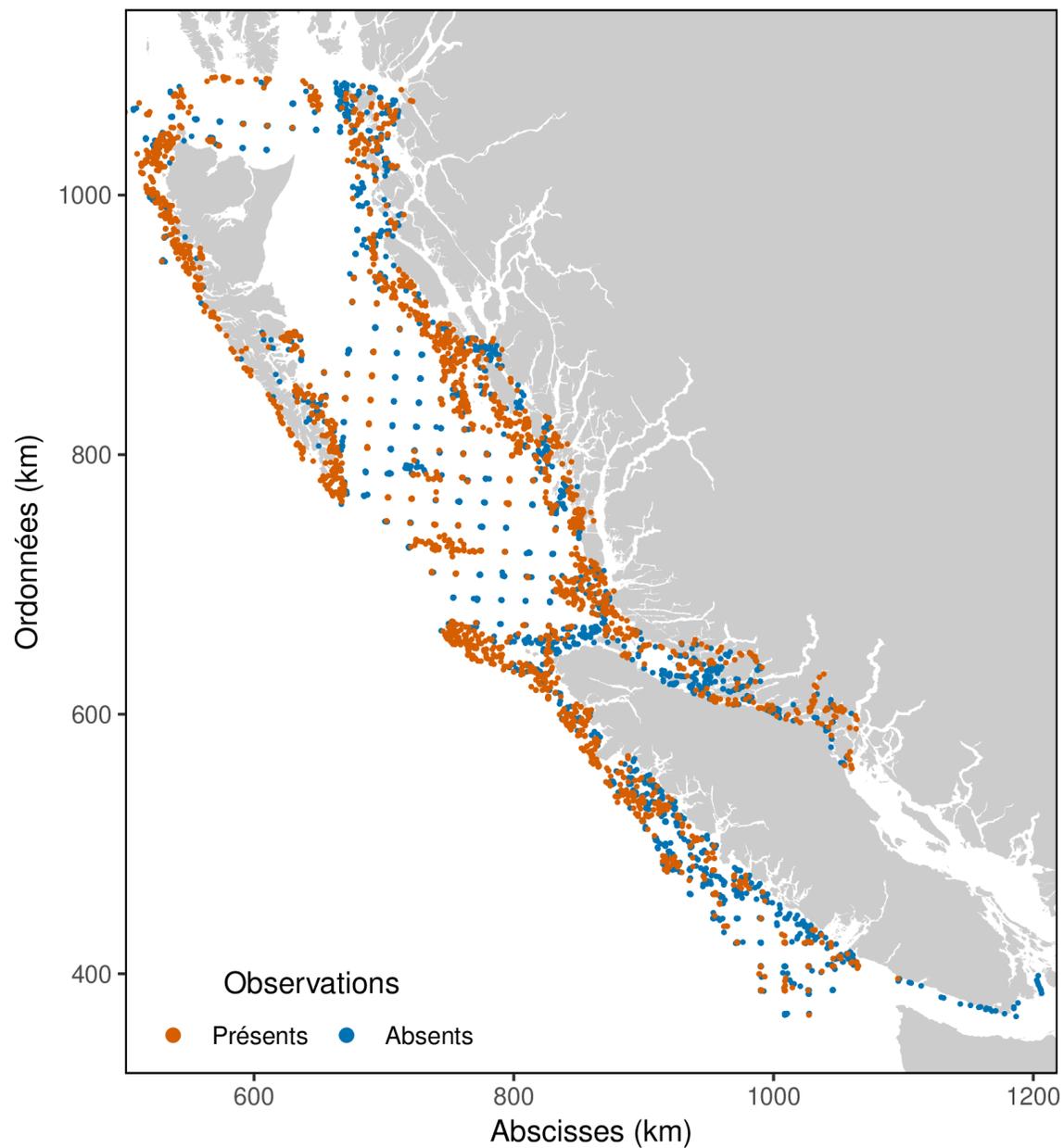


Figure A.19 Observations de la présence et de l'absence du sébaste aux yeux jaunes dans la zone d'étude du plateau.

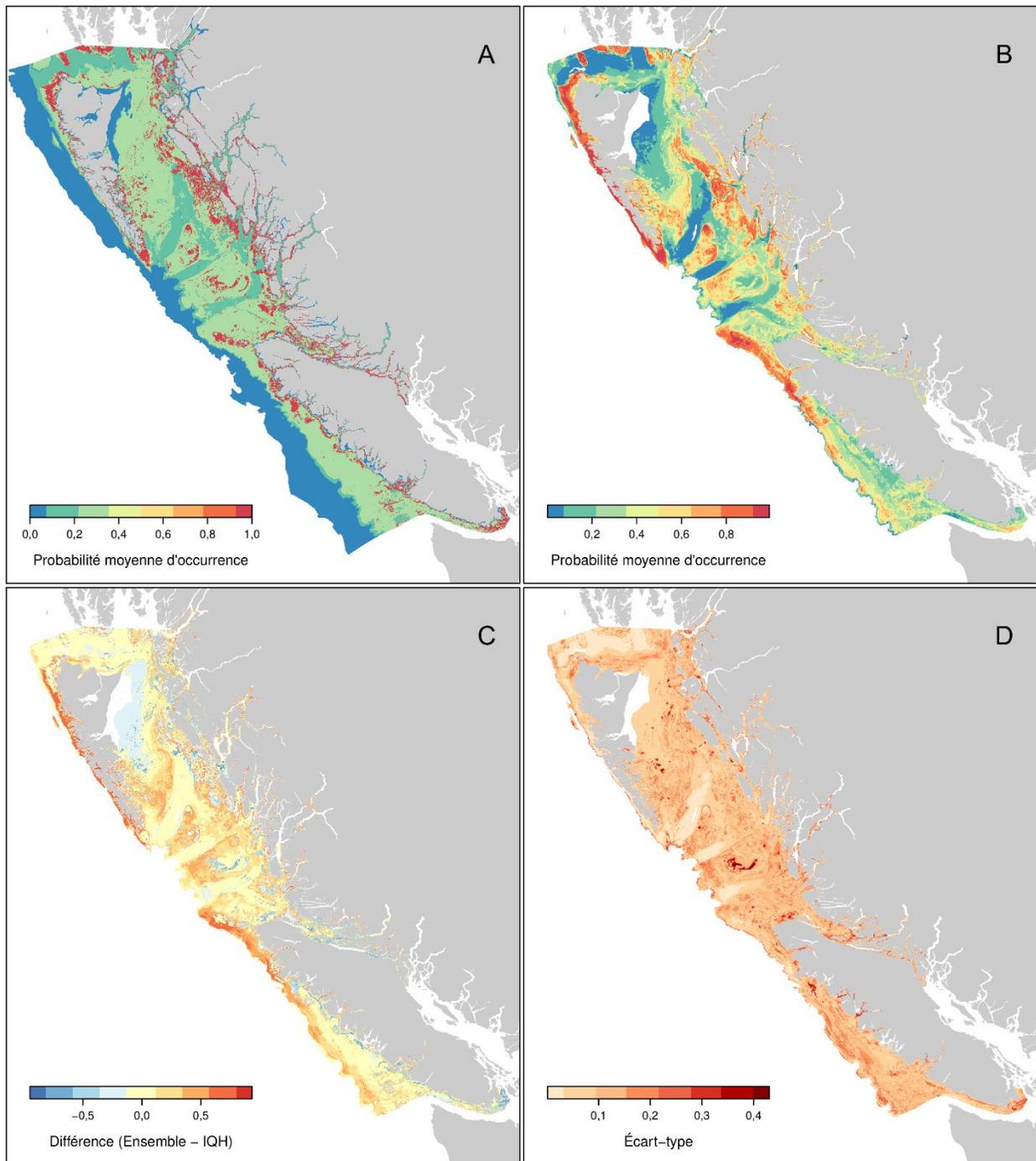


Figure A.20. Prédications de la répartition du sébaste aux yeux jaunes et incertitude correspondante. Prédications de la probabilité d'occurrence à partir A) du modèle d'indice de la qualité de l'habitat (IQH) et B) du modèle d'ensemble basé sur des modèles linéaires généralisés et d'arbres à régression stimulée. L'incertitude du modèle est représentée par C) la différence entre les prédictions du modèle IQH et celles du modèle d'ensemble et D) l'écart-type des prédictions de plusieurs modèles d'ensemble.

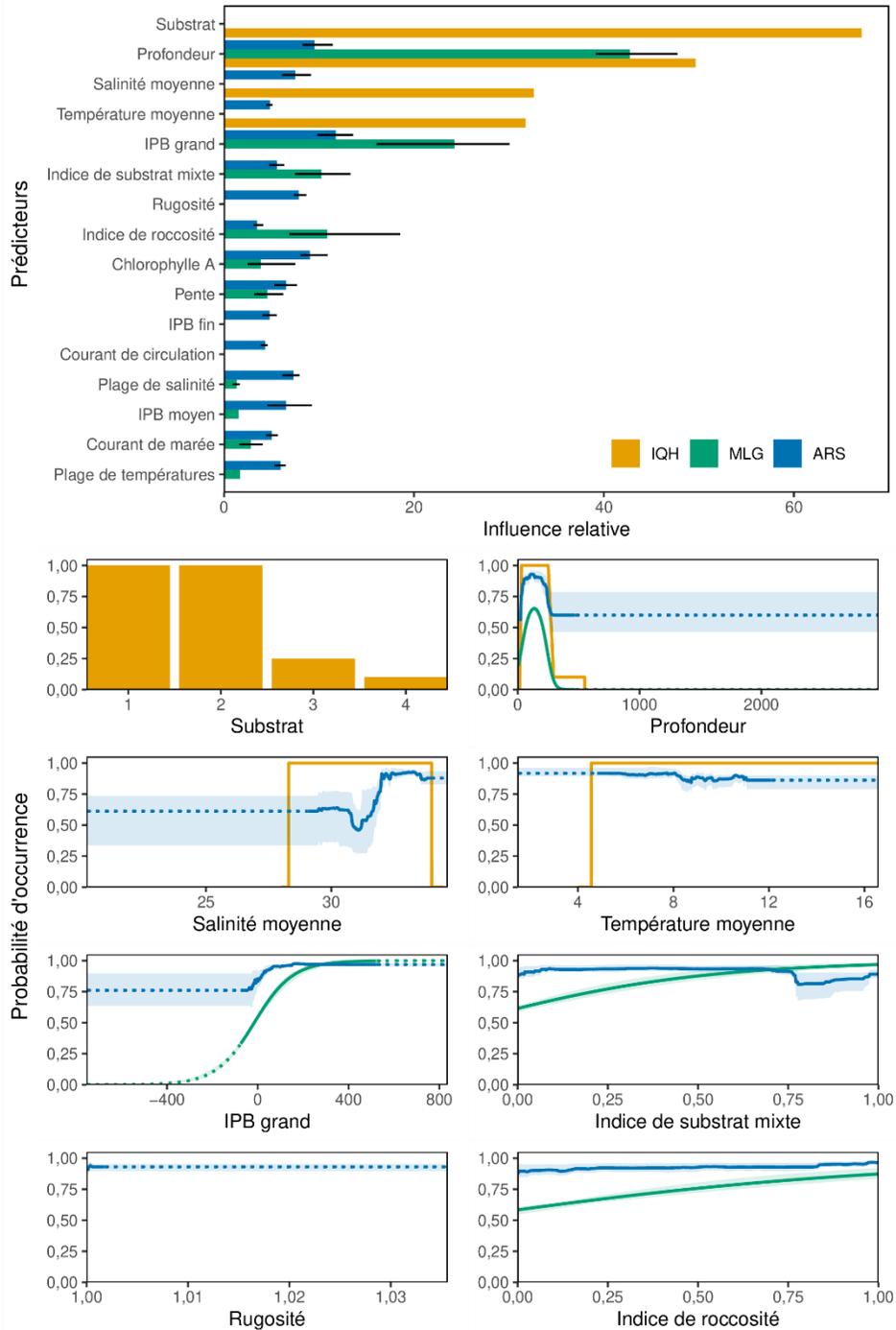


Figure A.21. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) pour les huit prédicteurs environnementaux les plus influents de chacun des modèles IQH, MLG et ARS du sébaste aux yeux jaunes. Pour les modèles MLG et ARS, les barres des diagrammes de l'influence relative représentent la moyenne, et les lignes montrent l'influence relative minimale et maximale des modèles à validation croisée à cinq plis. Dans les diagrammes des effets marginaux, les lignes pleines représentent les effets marginaux moyens par méthode, et les zones ombrées représentent les effets marginaux minimaux et maximaux des modèles à validation croisée à cinq plis. Le substrat a été représenté comme une variable catégorielle pour le modèle IQH et comme un indice de roccosité continu pour les modèles MLG et ARS.

Étoile ocrée

Pisaster ochraceus

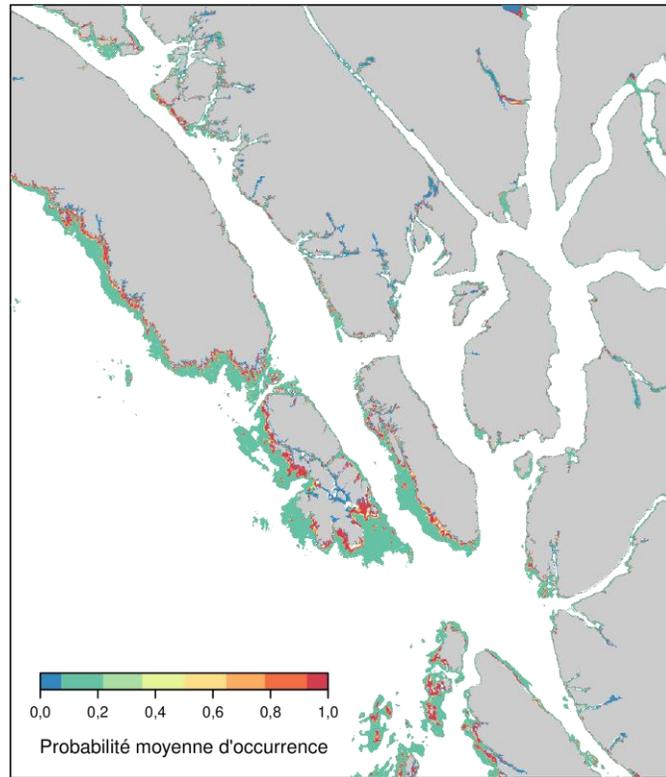


Figure A.22. Prédiction de la répartition de l'étoile ocrée à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée sur la figure A.1 par la case en pointillés.

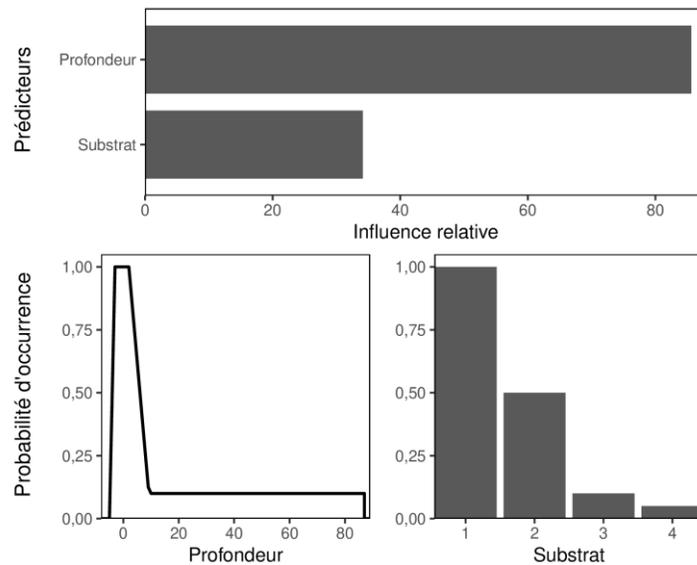


Figure A.23. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de l'étoile ocrée.

Complexe de moules bleues

Mytilus edulis, *M. trossulus* et *M. galloprovincialis*

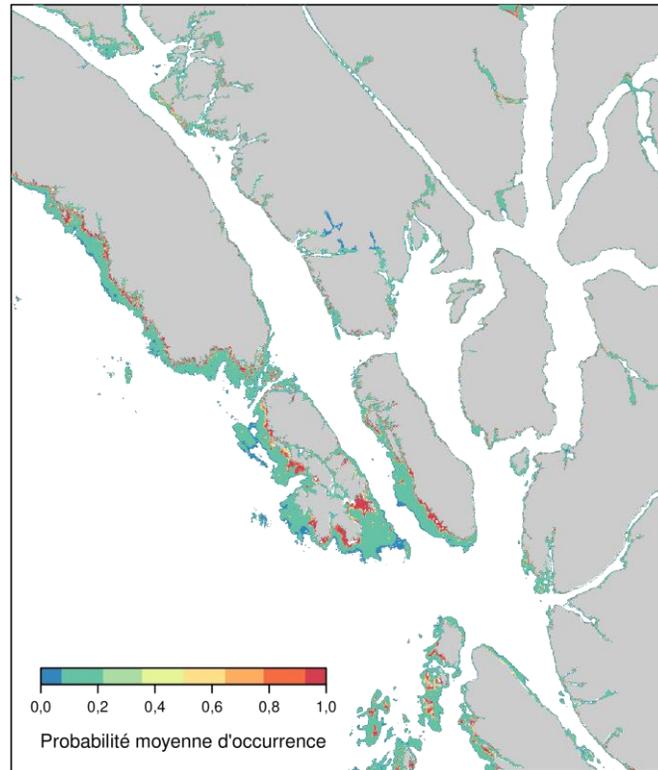


Figure A.24. Prédications de la répartition de la moule bleue à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.

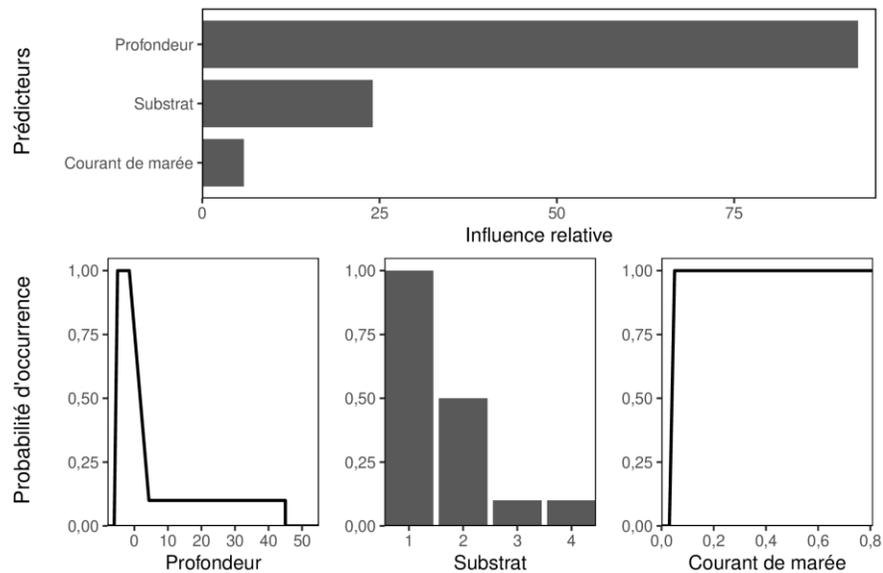


Figure A.25. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la moule bleue.

Palourde du Pacifique

Leukoma staminea

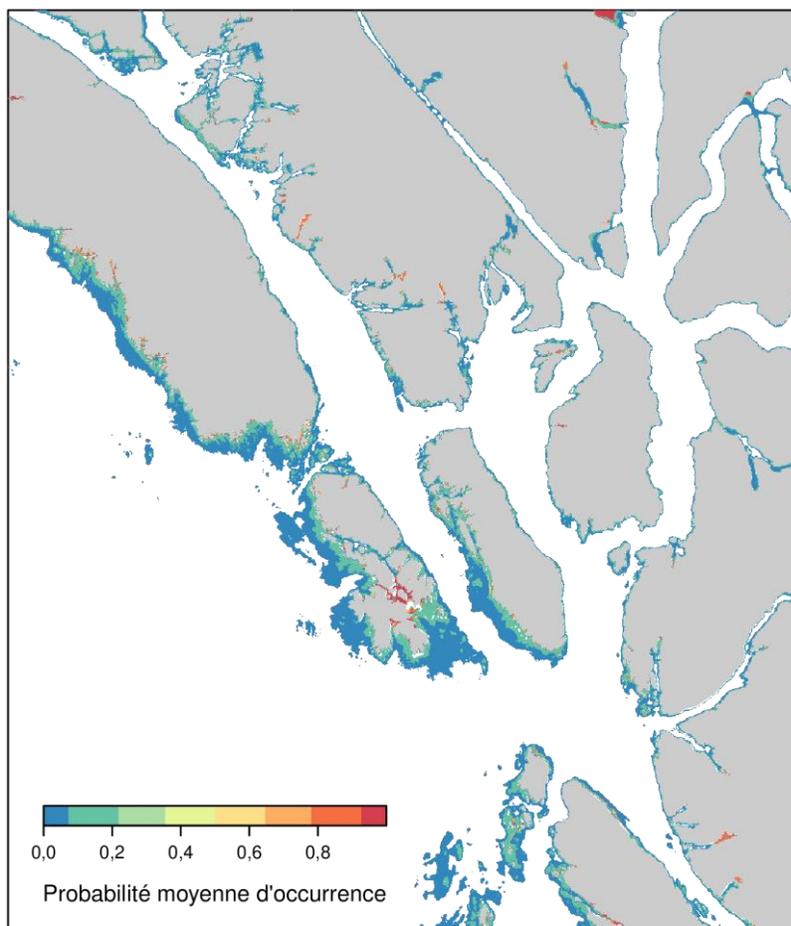


Figure A.26. Prédications de la répartition de la palourde du Pacifique à partir d'un modèle d'indice de qualité de l'habitat. La zone représentée ici est indiquée à la figure A.1 par la case en pointillés.

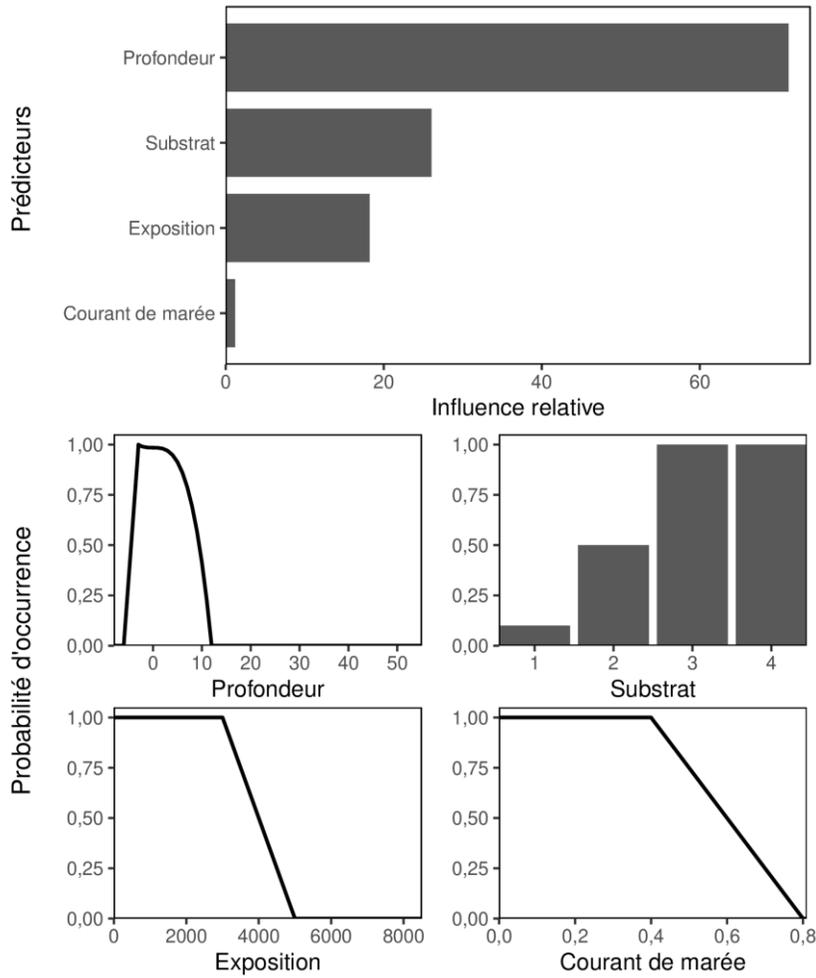


Figure A.27. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la palourde du Pacifique.

Pennatule orangée

Ptilosarcus gurneyi

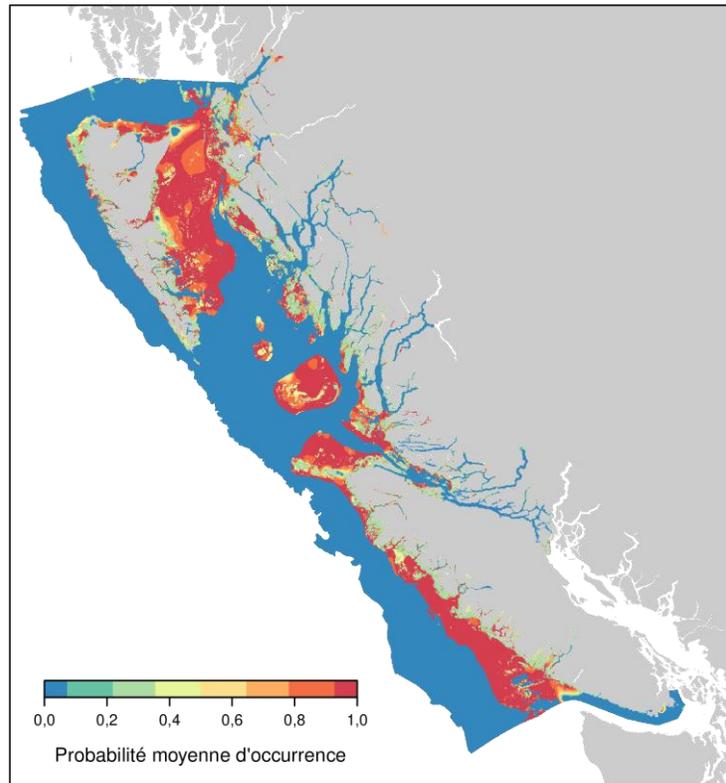


Figure A.28. Prédications de la répartition de la pennatule orangée à partir d'un modèle d'indice de qualité de l'habitat.

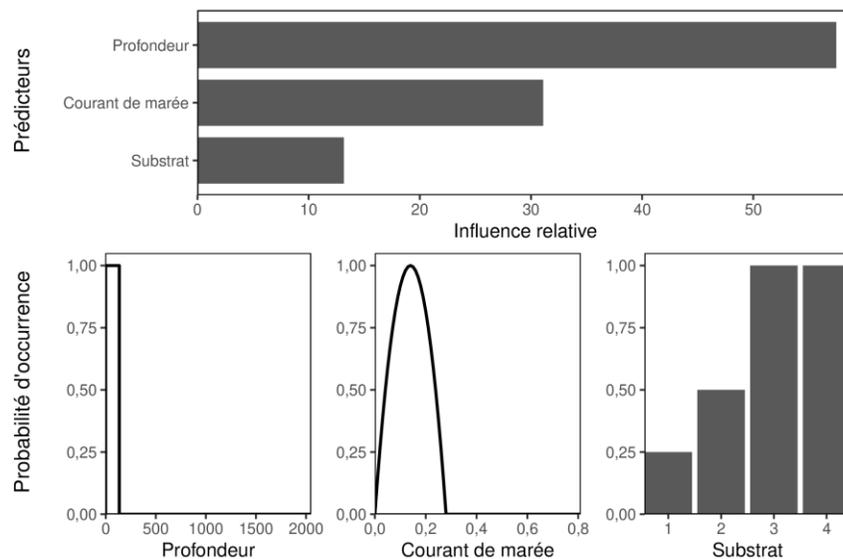


Figure A.29. Influence relative des prédicteurs (en haut) et des effets marginaux (en bas, plusieurs panneaux) d'après le modèle d'indice de qualité de l'habitat de la pennatule orangée.

ANNEXE B. DESCRIPTIONS DE L'INDICE DE QUALITÉ DE L'HABITAT

Les modèles d'indice de qualité de l'habitat (IQH) sont des modèles d'enveloppe qui s'appuient sur l'opinion d'experts pour poser des hypothèses permettant de relier la qualité de l'habitat aux facteurs abiotiques. Les méthodes de construction des modèles IQH sont détaillées à la section 4.2.5. Nous décrivons ici l'écologie pertinente et les hypothèses formulées décrivant les relations des prédicteurs environnementaux (abiotiques) prises en compte pour chaque espèce. Nous avons accordé la préférence aux connaissances portant spécialement sur la Colombie-Britannique (C.-B.). Lorsqu'il n'en existait pas, nous avons utilisé les connaissances sur l'ensemble de l'est du Pacifique Nord. Tous les prédicteurs sont censés être limitatifs, de sorte que l'indice de qualité à un endroit donné est égal à la valeur prédictive la plus faible.

Des relations prédictives ont été établies spécialement pour les couches de prédicteurs environnementaux suivantes (décrites à la section 4.2.2), le cas échéant : substrat (1 = roche, 2 = mélange, 3 = sable, 4 = vase); profondeur (m); salinité moyenne au fond en été (USP); température moyenne au fond en été (°C); vitesse moyenne des courants de marée en été ($m \cdot s^{-1}$); et exposition (en centaines de kilomètres). La relation entre les variables environnementales et la qualité de l'habitat peut changer en fonction des échelles temporelles et spatiales de ces prédicteurs. Ainsi, la construction d'enveloppes d'habitat dérivées d'avis spécialisés ne tient pas toujours compte des processus écologiques à petite échelle. Au fur et à mesure que des couches de prédicteurs à plus haute résolution deviennent disponibles, il faudra réévaluer les relations décrites dans le présent document et les modifier au besoin.

Au cours du processus de modélisation, nous avons constaté que la couche de substrat était biaisée en faveur du substrat rocheux (c.-à-d. que les experts ont remarqué que la roche était souvent signalée dans des zones connues pour leur substrat meuble). Dans une moindre mesure, il y a aussi des prédictions de substrat meuble dans des zones connues pour être rocheuses. Pour gérer cette incertitude, nous avons attribué des catégories de substrat allant de qualité marginale à inadéquate (0,05 à 0,25 au lieu de 0) afin que les habitats potentiellement convenables (mais mal caractérisés) ne soient pas désignés comme totalement inadéquats.

Ormeau nordique

Haliotis kamtschatkana

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité, la vitesse des courants de marée et l'exposition dans le modèle IQH pour l'ormeau nordique. Il n'est pas possible d'afficher les prédictions du modèle de répartition de l'ormeau nordique en raison de son statut d'espèce en péril. Toutefois, l'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS; figure 4.3) montre que ce sont l'exposition et la profondeur qui ont eu la plus grande influence sur la prédiction du modèle IQH.

L'ormeau nordique est généralement présent de la zone intertidale basse à environ 100 m de profondeur, le plus souvent à moins de 10 m de profondeur dans les eaux de la Colombie-Britannique (Sloan et Breen, 1988). Nous avons attribué la qualité comme augmentant de 0 à 1 entre -6 m de profondeur (ligne des hautes eaux) et -3 m (moitié de la zone intertidale) pour permettre une transition entre un habitat inadéquat et un habitat pleinement convenable. La qualité est demeurée à 1 jusqu'à 10 m, profondeur à partir de laquelle nous avons appliqué une diminution accélérée pour tenir compte du fait que la majeure partie de la population se trouve à moins de 10 m de profondeur. Nous avons assigné une qualité de 0 après 100 m, la limite déclarée de la répartition de l'ormeau nordique en profondeur.

L'ormeau nordique se trouve habituellement sur des substrats fermes comme les roches, les blocs rocheux et la couche rocheuse (MPO, 2007), et l'habitat optimal est composé de diverses combinaisons de rebords, de replis et de dépressions dans les pierres ainsi que de blocs rocheux empilés (Emmett et Jamieson, 1988). Les ormeaux nordiques juvéniles (longueur de coquille de 10 à 70 mm) se trouvent sous et sur les zones exposées des roches (MPO, 2007) et sont habituellement associés aux algues encroûtantes (Sloan et Breen, 1988). Les adultes (longueur de coquille > 70 mm) vivent sur des surfaces rocheuses exposées (MPO, 2007) et colonisent souvent les peuplements de varech (Williams, 1989). Nous avons attribué une qualité de 1 pour les roches, de 0,5 pour les substrats mixtes et de 0,1 pour les substrats de sable et de vase.

L'ormeau nordique a besoin d'eau de mer à salinité totale (> 30 ppt) (COSEWIC, 2009). Nous avons attribué la salinité comme étant inadéquate aux valeurs < 28 USP (équivalent aux ppt), avec une augmentation linéaire de la qualité entre 28 et 30 USP pour permettre une transition entre les salinités inadéquates et les salinités pleinement convenables, avec une pleine qualité de 1 au-delà.

Lorsqu'il a été exposé à des températures entre 2 °C et 24 °C en laboratoire, l'ormeau nordique n'a présenté aucun signe de mortalité ou de stress respiratoire (Paul et Paul, 1998). La plage des températures moyennes au fond en été dans notre zone d'étude se situe entièrement dans la plage de qualité établie. Nous n'avons donc pas inclus la température comme variable explicative.

L'ormeau nordique est présent dans une vaste gamme d'habitats, allant de baies assez abritées à des côtes exposées (COSEWIC, 2009), et se trouve habituellement dans des zones d'échange d'eau de mer modéré à élevé, comme les côtes exposées ou semi-exposées (MPO, 2007). La croissance est plus rapide dans les zones modérément exposées que dans les zones très exposées, car la forte action des vagues et les courants réduisent les possibilités pour l'ormeau de « capturer » les algues dérivantes dont il se nourrit (Sloan et Breen, 1988). Des études menées dans le bassin de Barkley et aux îles de la Reine-Charlotte ont révélé que le transfert de l'ormeau nordique d'un habitat exposé à un habitat plus abrité augmente les taux de croissance (Breen, 1986; Emmett et Jamieson, 1988), tandis que plusieurs études réalisées en Colombie-Britannique ont déterminé que la densité de l'espèce augmente avec l'exposition

relative aux vagues, mais que la longueur moyenne des coquilles diminue (Breen et Adkins, 1979; Lessard et Campbell, 2007; Tomascik et Holmes, 2003). Puisque les taux de croissance augmentent dans les zones plus abritées, les densités plus faibles peuvent être le résultat de pêcheurs et de braconniers historiques qui ont ciblé des zones plus faciles d'accès et qui n'ont pas réduit la qualité de l'habitat, mais nous n'avons pas été en mesure de trouver de preuves à l'appui de cette hypothèse. Nous avons supposé que l'exposition était au moins quelque peu convenable sur toute la gamme des valeurs dans la zone d'étude (0,005 à 84,9 *100 km) et nous avons attribué une qualité optimale de 1 pour les valeurs d'exposition moyennes à élevées entre 30 et 70 (en centaines de kilomètres). Nous avons attribué une relation croissante entre 0 et 30 et une relation décroissante entre 70 et 84,9 pour tenir compte du fait que la capacité de se nourrir de varech dérivant est probablement réduite lorsque l'action des vagues est faible et élevée. Il convient toutefois de noter que la relation est ambiguë lorsque l'exposition est faible.

Bien que l'influence des courants de marée sur l'ormeau nordique ne soit pas bien étudiée, Lessard et Campbell (2007) indiquent que l'espèce a besoin d'un bon échange d'eau, avec un courant de marée ou une action des vagues. Nous avons supposé que les possibilités de capturer des algues dérivantes pour s'en nourrir augmenteraient avec des vitesses de courant plus élevées, et nous avons attribué une relation linéaire croissante avec la vitesse du courant de marée sur l'ensemble de la plage relevée dans la zone d'étude. Il est probable que la capture des algues dérivantes serait réduite à des vitesses de courant plus élevées, mais la plage des valeurs de la vitesse moyenne des courants de marée dans la zone d'étude a une valeur maximale de 0,81 m·s⁻¹ (1,57 nœud), ce qui est encore relativement faible; nous avons supposé que l'alimentation ne serait pas inhibée à cette vitesse.

Étant donné la préférence pour les habitats présentant une certaine complexité structurale, en particulier pour les juvéniles qui s'abritent sous les roches, nous avons tenté d'inclure un prédicteur de la complexité de l'habitat, en créant d'abord un indice de qualité pour la rugosité, puis un pour la pente. Les indices ainsi obtenus étaient très restrictifs, produisant des modèles IQH avec très peu d'habitats convenables. Nous avons présumé que la résolution des données prédictives n'était pas assez élevée pour refléter les processus agissant sur la répartition de l'ormeau nordique. C'est pourquoi nous n'avons pas inclus la complexité de l'habitat dans le modèle.

Panope du Pacifique

Panopea generosa

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité, la vitesse des courants de marée et l'exposition dans le modèle d'enveloppe IQH pour la panope (figure A.2). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que ce sont le substrat, la profondeur et l'exposition qui ont eu la plus grande influence sur la prédiction du modèle IQH (figure A.3).

Nous avons fondé la distribution de profondeur préférée sur les données de Campbell *et al.* (1998) et sur les données des relevés en plongée sur la densité recueillies dans la zone d'étude. Campbell *et al.* (1998) indiquent que les densités de la panope augmentent entre des profondeurs de 1 m à environ 20 ou 24 m, puis diminuent jusqu'à 110 m. Cela correspond aux données de la zone d'étude, qui montrent un pic de densité entre 10 et 17 m, bien que les données soient limitées aux profondeurs de moins de 20 m (Bureau¹). Cependant, on trouve aussi des gisements de panopes dans les zones intertidales offrant des substrats et des conditions convenables. Nous avons donc construit une préférence de profondeur optimale avec une augmentation linéaire de la qualité de 0 à 1 entre -2 et 0 m de profondeur, une pleine qualité jusqu'à 25 m, puis une diminution linéaire de la qualité jusqu'à 0 à 110 m (ce qui donne une valeur de 0,7 à 50 m de profondeur, la limite de la zone d'étude).

En tant qu'organismes benthiques, les panopes sont principalement associées aux habitats à fond meuble, peuvent se trouver sur des substrats mixtes et sont absentes sur les fonds durs (Goodwin et Pease, 1989). Nous avons attribué une qualité de 1 aux substrats de sable et de vase, de 0,5 aux substrats mixtes et de 0,25 à la roche. Bien que les panopes ne vivent pas sur les fonds durs, la valeur attribuée de 0,25 à la roche explique une surprédiction de la roche dans la couche du substrat, certaines zones connues pour leur substrat meuble étant prédites comme ayant un substrat rocheux.

Goodwin (1973) a déterminé les plages optimales de salinités et de températures au fond pour la survie des larves. Nous avons repris ces valeurs pour définir une courbe de qualité optimale, mais comme la température moyenne estivale au fond observée dans la zone d'étude se situait presque entièrement à l'intérieur des valeurs optimales, ce prédicteur n'a pas été inclus dans le modèle.

Une énergie élevée des vagues de vent peut entraîner une mauvaise stabilité du substrat, ce qui réduit la qualité de l'habitat pour la panope. Cependant, les panopes semblent préférer un certain mouvement de l'eau, car on ne les trouve pas dans des endroits très abrités, comme les têtes des bras de mer (Bureau¹). Nous avons donc exclu les zones d'exposition très faible et modérée à élevée en attribuant une augmentation linéaire de la qualité de 0 à 1 entre les valeurs d'exposition de 0 et 1 (en centaines de kilomètres), une qualité optimale de 1 à 25, puis une diminution linéaire jusqu'à 0 à 50. Nous avons également pris en compte un certain mouvement de l'eau sous l'effet des marées, en excluant les zones où la vitesse maximale des courants de marée est inférieure à $3 \text{ cm}\cdot\text{s}^{-1}$, puis en augmentant la qualité de l'habitat jusqu'à un maximum à $5 \text{ cm}\cdot\text{s}^{-1}$. Les valeurs d'exposition et de marée ont été établies en consultation avec D. Bureau en fonction de la position des gisements de panopes connus.

¹ Bureau, D. 2018. Biologiste, Division des ressources aquatiques, de la recherche et de l'évaluation (DRERA), Mollusques et crustacés, Sciences, Pêches et Océans Canada. Mars 2015 et décembre 2018, communication personnelle.

Oursin rouge

Mesocentrotus franciscanus

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur (conditionnée par l'exposition) et la salinité dans le modèle d'enveloppe IQH pour l'oursin rouge (figure A.5). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle IQH (figure A.6).

L'oursin occupe principalement des substrats rocheux, bien qu'il puisse se trouver sur des substrats mixtes ou meubles dans les zones à faible énergie. Nous avons supposé une pleine qualité de 1 sur le substrat rocheux, une qualité de 0,5 sur un substrat mixte, de 0,1 sur le sable et de 0,05 sur la vase.

L'oursin rouge est généralement réparti de la zone intertidale jusqu'à une profondeur de 50 m (DFO, 2018), bien qu'on puisse trouver des individus jusqu'à 284 m (Galloway²). Dan Leus³ a souligné l'importance de l'interaction entre l'énergie et la profondeur, ce qui a conduit à distinguer les préférences en matière de profondeur en fonction du régime énergétique. Pour créer l'interaction profondeur-exposition, nous avons attribué une relation de profondeur à faible exposition dans les zones où les valeurs de l'exposition étaient de 20 (en centaines de kilomètres) ou moins et une relation de profondeur à forte exposition dans les zones où les valeurs de l'exposition étaient supérieures à 20 (en centaines de kilomètres). Ces relations ont été combinées en un seul prédicteur de la profondeur, conditionné par l'exposition. Pour la relation de profondeur à faible exposition, nous avons attribué une profondeur convenable dans la plage de -6 à 284 m, avec une augmentation linéaire de la qualité de 0 à 1 entre -6 et -4 m de profondeur, une qualité optimale de 1 entre -4 et 9 m de profondeur, puis une diminution linéaire jusqu'à une qualité marginale de 0,1 à 14 m de profondeur. Pour la relation de profondeur à forte exposition, nous avons attribué une profondeur convenable dans la plage de -3 à 284 m, avec une augmentation linéaire de la qualité de 0 à 1 entre 0 et 2 m de profondeur, une qualité optimale de 1 entre 2 et 15 m de profondeur, puis une diminution linéaire jusqu'à une qualité marginale de 0,1 à 20 m de profondeur. Pour la relation de profondeur à forte exposition, les valeurs supérieures de la plage ont été décalées vers le bas, car les oursins se déplaceront plus profondément pour éviter d'être emportés par les eaux dans les zones fortement exposées. Nous avons supposé que cet effet ne serait pas observé en dessous de 20 m de profondeur. Dans les relations de profondeur à exposition forte et faible, la qualité augmente linéairement de 0 à 1 au sommet de la plage de profondeurs pour refléter la transition entre les profondeurs inadéquates et les profondeurs convenables. De même, il y a une transition d'une profondeur entièrement convenable à une profondeur marginalement convenable juste en dessous de la plage de profondeurs optimale dans les deux relations.

La salinité a été attribuée comme étant inadéquate en dessous de 28 USP, avec une augmentation linéaire de la qualité entre 28 et 31 USP et une pleine qualité au-delà.

Ces oursins préfèrent les courants modérés à forts (DFO, 2018), car il y aura plus de nourriture disponible dans ces conditions, mais évitent les zones très exposées où la pression de l'eau

² Galloway. Professeur adjoint, Université de l'Oregon, Oregon Institute of Marine Biology. Octobre 2018, communication personnelle.

³ Leus, D. Biologiste spécialiste des oursins, Section des invertébrés marins, Division de l'évaluation des stocks et de la recherche, Sciences, Pêches et Océans Canada. 2015 et décembre 2018, communication personnelle.

pourrait les retourner et les exposer à la prédation (Leus³). Nous avons inclus l'exposition comme condition pour la profondeur, une exposition forte amenant les oursins à se déplacer plus profondément pour éviter la forte action des vagues. La vitesse des courants de marée n'a pas été incluse dans le modèle, car les prédicteurs n'étaient pas à une échelle appropriée pour résoudre les processus écologiques présents.

Varech *pterygophora*

Pterygophora californica

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité, la vitesse des courants de marée et l'exposition dans le modèle d'enveloppe IQH pour le varech *Pterygophora* (figure A.8). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle IQH (figure A.9).

Le varech *Pterygophora* est abondant sur les roches et les galets (Druehl et Clarkston, 2016; Hawkes *et al.*, 1979) et nous avons supposé une pleine qualité de 1 sur le substrat rocheux, une qualité de 0,5 sur le substrat mixte et une qualité 0,1 sur le sable et la vase.

Cette espèce se trouve dans la zone infratidale, entre 2 et 20 m de profondeur (Mondragon et Mondragon, 2003). Nous avons attribué une profondeur convenable entre 0 m (laisse de basse mer) et 20 m, avec une augmentation linéaire de la qualité de 0 à 1 entre 0 et 2 m de profondeur, une pleine qualité de 1 entre 2 et 10 m, puis une diminution retardée jusqu'à 0 à 20 m. L'augmentation linéaire de 0 à 2 m a été incluse pour permettre une transition entre un habitat pleinement convenable et un habitat non approprié ou marginalement convenable.

Au cours de l'examen préliminaire, il a été suggéré d'utiliser les valeurs de tolérance à la salinité de l'ormeau nordique, qui est présent dans un habitat semblable à celui du varech *Pterygophora*. L'ormeau nordique préfère l'eau de mer à salinité totale (> 30 USP) (COSEWIC, 2009). Nous avons attribué une salinité non convenable à 28 USP, avec une qualité croissante entre 28 et 30 USP et une pleine qualité au-delà. La qualité croissante entre 28 et 30 USP a été prise en compte pour créer une transition d'un habitat inadéquat à un habitat pleinement convenable.

Cette espèce s'établit généralement sur des rochers dans les zones exposées aux vagues (Druehl et Clarkston, 2016) et est adaptée aux habitats balayés par les vagues où les courants sont forts (Lamb et Hanby, 2006). L'exposition a été jugée appropriée entre 0 et 84,9 (en centaines de kilomètres), avec une augmentation linéaire de la qualité de 0 à 1 entre 0 et 0,3, une pleine qualité de 1 entre 0,3 et 30, puis une diminution linéaire jusqu'à 0 de la qualité à 84,9 (en centaines de kilomètres).

La vitesse du courant de marée a été assignée comme étant convenable de 0 à 0,81 m·s⁻¹, avec une distribution normale sur toute cette plage, ce qui donne une pleine qualité de 1 à 0,4 m·s⁻¹.

Zostère marine

Zostera spp.

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité, la température, la vitesse des courants de marée et l'exposition dans le modèle IQH pour la zostère marine (figure A.11). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est la profondeur et le substrat qui ont eu la plus grande influence sur la prédiction du modèle IQH (figure A.12).

La zostère marine est présente à partir du milieu de la zone intertidale jusqu'à 30 m de profondeur (Phillips et Lewis, 1983). En consultation avec des experts, la profondeur a été assignée comme convenable entre -6 m (laisse des hautes eaux) à 30 m, avec une augmentation linéaire de la qualité de 0 à 1 entre -6 et -3,3 m de profondeur, une pleine qualité de 1 entre -3,3 et 10 m, une diminution linéaire à une qualité marginale de 0,1 à 20 m, puis une qualité de 0 au-delà de 30 m de profondeur. L'augmentation et la diminution linéaires de la qualité entre -6 et 3,3 m et entre 10 et 20 m (respectivement) ont été incluses pour permettre des transitions entre des habitats totalement convenables et inadéquats ou marginalement convenables.

Cette espèce est enracinée dans des sédiments sablonneux ou boueux (Vandermeulen, 2005), l'habitat optimal étant composé d'un mélange de sable et de vase (Phillips et Lewis, 1983). Pour le substrat, nous avons supposé une pleine qualité de 1 sur le substrat sablonneux, et une qualité de 0,75 sur la vase, de 0,5 sur un mélange et de 0,25 sur un substrat rocheux.

La zostère marine peut tolérer une plage étendue de températures, allant de -6 à 40,5 °C, la plage optimale étant de 10 à 20 °C (Phillips et Lewis, 1983). La zostère marine peut également tolérer un large éventail de salinités, de l'eau douce à 42 USP, avec une salinité optimale entre 10 et 30 USP (Phillips et Lewis, 1983). Nous avons utilisé les plages et les valeurs optimales indiquées dans la documentation pour la température et la salinité. Une pleine qualité de 1 a été attribuée pour l'ensemble de la plage optimale, avec une diminution linéaire de la qualité jusqu'à 0 aux limites de la plage complète pour chaque prédicteur.

L'habitat optimal est soumis à une légère action des vagues (Phillips et Lewis, 1983), et la zostère marine se rencontre couramment dans les eaux protégées des baies et des estuaires (Mondragon et Mondragon, 2003). Pour la qualité de l'exposition, nous avons attribué une relation linéaire décroissante de 0 à 60 (en centaines de kilomètres), avec une qualité de 0 au-delà. Cette relation a conduit à désigner les zones à forte exposition comme non convenables, car l'habitat optimal de la zostère marine est caractérisé par une légère action des vagues.

La vitesse du courant de marée a été assignée comme étant convenable de 0 à 0,81 m·s⁻¹, avec une distribution normale sur toute cette plage, ce qui donne une pleine qualité de 1 à 0,4 m·s⁻¹. Cette relation permet d'attribuer une qualité marginale aux habitats stagnants et aux habitats où le courant est fort.

Crabe dormeur

Metacarcinus magister

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité et la température dans le modèle IQH pour le crabe dormeur (figure A.14). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle IQH (figure A.15). Les autres prédicteurs étaient cependant très influents également. En effet, tous les prédicteurs présentaient une pleine qualité sur une grande partie de la zone d'étude.

Le crabe dormeur est présent de la zone intertidale basse jusqu'à 230 m de profondeur (DFO, 2017). Dans leur étude d'un estuaire de l'Alaska, Stone et O'Clair (2001) font état de préférences de profondeur variables selon le sexe et la saison, allant de moins de 10 m à plus de 70 m, ce qui nous a amenés à supposer une qualité de 1 sur toute la plage de profondeurs pour le crabe dormeur (-2 à 230 m).

Le crabe dormeur habite habituellement des substrats composés de sable, de vase ou de limon (DFO, 2017), bien qu'il ait été observé sur des substrats durs (J. Lessard, observation personnelle). Compte tenu de ces préférences, nous attribuons une qualité de 1 au fond meuble, de 0,5 au fond mixte et de 0,1 au fond dur.

Bien qu'une grande partie de leur habitat se trouve dans des zones estuariennes exposées à des épisodes fréquents de faible salinité, les crabes dormeurs sont considérés comme des osmorégulateurs faibles (Engelhardt et Dehnel, 1973). Ils présentent des signes comportementaux et physiologiques de stress à des salinités inférieures à 24 ppt (Curtis et McGaw, 2012; McGaw *et al.*, 1999) et sont incapables de tolérer une exposition prolongée à moins de 12 ppt (Cleaver, 1949). Compte tenu des données de terrain, où les crabes dormeurs ont rarement été observés à des salinités inférieures à 10 ppt (Curtis et McGaw, 2008), nous attribuons une qualité croissante de 0 à 1 entre 10 et 24 USP (équivalent aux ppt) et une pleine qualité de 1 au-delà.

Pour la température de l'eau, la croissance et la survie sont limitées au-dessus de 15 °C en raison de l'augmentation des dépenses énergétiques associées à la respiration (Kondzela et Shirley, 1993). La limite supérieure de température létale pour le crabe dormeur varie de 20 à 25 °C, selon la saison (Pauley *et al.*, 1989). Dans une expérience en laboratoire où Curtis et McGaw (2012) ont mis le crabe dormeur à l'épreuve en le soumettant à des stress de famine, de températures et de salinités, les crabes ont toléré des températures maximales de 15 °C. Nous attribuons une qualité optimale de 1 à des températures allant jusqu'à 15 °C, la qualité diminuant jusqu'à 0 à 25 °C.

Le crabe dormeur est un prédateur actif des mollusques et crustacés, généralement des palourdes, et peut aussi attraper des poissons. Il a une forte préférence pour les aliments frais, mais se nourrira aussi de carcasses de proies mortes (Curtis, 2019⁴). La prédation exercée par le crabe dormeur dans les estuaires est un exercice d'équilibre entre les éprouves physiologiques et la disponibilité des palourdes en tant que proies, qui sont habituellement présentes dans des eaux moins profondes et moins salées. Cela signifie que, à une échelle plus régionale, les estuaires préférés des palourdes constitueraient un habitat important pour les crabes dormeurs. Pour le crabe, l'habitat estuarien est une fonction de la salinité, de la

⁴ Curtis, D. Biologiste, Section des mollusques, Division de l'évaluation des stocks et de la recherche, Pêches et Océans Canada. Janvier 2019, communication personnelle.

température et de l'exposition, car l'accessibilité des palourdes en tant que proies dépend des eaux stratifiées qui permettent la formation d'un coin salin pendant la marée montante (Curtis, 2019⁴). Il n'existe actuellement aucun ensemble de données sur l'exposition à la bonne résolution spatiale pour la zone d'étude du plateau; nous avons donc tenté d'utiliser la vitesse des courants de marée pour mettre en évidence les zones de faible mouvement des eaux où ce coin salin peut se former. Nous avons constaté que la résolution de 100 m sur 100 m utilisée dans cette zone d'étude était trop grossière pour résoudre les processus à échelle fine comme la formation de coins de salinité. C'est pourquoi la vitesse du courant de marée n'a pas été incluse dans le modèle.

Sébaste à dos épineux

Sebastes maliger

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité et la température dans le modèle IQH pour le sébaste à dos épineux (figure A.17). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est le substrat qui a eu la plus grande influence sur la prédiction du modèle IQH (figure A.18). Les autres prédicteurs étaient cependant très influents également. En effet, tous les prédicteurs présentaient une pleine qualité sur une grande partie de la zone d'étude.

Au cours de l'examen du modèle préliminaire, il a été suggéré de créer des modèles distincts pour les juvéniles et les adultes. Toutefois, compte tenu des contraintes de temps, nous avons modélisé les juvéniles et les adultes ensemble et nous créerons des modèles distincts à une date ultérieure.

La répartition du sébaste à dos épineux s'étend depuis la péninsule Kenai, dans le golfe d'Alaska, jusqu'au passage Anacapa, dans le sud de la Californie (Love *et al.*, 2002), et la profondeur est généralement considérée comme le plus important déterminant de la répartition (National Marine Fisheries Service, 2014). Dans toute son aire de répartition, cette espèce a été signalée dans la zone infratidale jusqu'à 274 m de profondeur (Love *et al.*, 2002). En Colombie-Britannique, les observations de sébastes à dos épineux varient de 9 à 186 m de profondeur (98 % de toutes les données de relevé de la base de données biologiques sur les poissons de fond – à l'exclusion des 1 % supérieur et inférieur). Par ailleurs, des juvéniles nouvellement établis ont probablement été observés dans les herbiers de zostère marine et d'autres zones intertidales (les juvéniles du sébaste cuivré, à dos épineux et brun sont impossibles à distinguer; CoppHaggarty, 2018⁵). Compte tenu de toutes ces aires de répartition signalées, nous avons attribué une pleine qualité de 1 pour une profondeur de 0 à 168 m, qui diminue jusqu'à 0 à une profondeur de 274 m.

Les sébastes à dos épineux se regroupent sur des substrats durs et complexes offrant un certain relief vertical, comme les roches brisées, les récifs rocheux, les crêtes et les crevasses (Richards, 1986) et les juvéniles peuvent aussi être observés sur des galets et des substrats meubles, comme les herbiers de zostère marine (Richards et Hand, 1987). Nous attribuons donc une pleine qualité de 1 pour les substrats rocheux et mixtes, de 0,5 pour le sable et une qualité marginale de 0,1 pour la vase.

En Colombie-Britannique, les sébastes à dos épineux sont observés durant les relevés au chalut de fond à des températures variant de 5,78 à 23 °C et à des salinités (relevés au chalut de fond de 2013 à 2018 et relevé à la palangre de 2018) comprises entre 28,2 et 33,9 USP (Keppel⁶). Les plages de températures et de salinités indiquées pour le sébaste à dos épineux et le sébaste aux yeux jaunes dans le sud-est de l'Alaska (Johnson *et al.*, 2003), près de l'extrémité nord de leur aire de répartition, se situent dans les plages de températures observées en Colombie-Britannique. Nous n'avons pas pu trouver les plages de températures ou de salinités signalées près de la partie sud de leur aire de répartition. Étant donné le manque d'information sur les préférences et les valeurs optimales, nous attribuons une pleine qualité de

⁵ Haggarty, D. Chef du Programme du sébaste et de la morue-langue dans les zones côtières, Section du poisson de fond, Division de l'évaluation des stocks et de la recherche, Pêches et Océans Canada. Décembre 2018, communication personnelle.

⁶ Keppel, E. Biologiste, Section du poisson de fond, Division de l'évaluation des stocks et de la recherche, Pêches et Océans Canada. Décembre 2018, communication personnelle.

1 à l'ensemble des plages moyennes de températures et de salinités observées pour la Colombie-Britannique.

Le sébaste à dos épineux préfère les habitats complexes comme la couche rocheuse fracturée et les complexes de blocs rocheux (p. ex. National Marine Fisheries Service, 2014; Yamanaka *et al.*, 2006). Pour exprimer cette complexité de l'habitat, nous avons d'abord créé un indice de qualité pour la rugosité, puis un indice pour la pente. Les indices ainsi obtenus étaient très restrictifs, produisant des modèles IQH avec très peu d'habitats convenables, car les données n'étaient pas à une échelle suffisamment fine pour refléter les processus agissant sur l'aire de répartition du sébaste à dos épineux. Nous avons donc choisi de ne pas inclure les indices de rugosité et de pente dans le modèle.

Sébaste aux yeux jaunes

Sebastes ruberrimus

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la salinité et la température dans le modèle IQH pour le sébaste aux yeux jaunes (figure A.20). L'influence relative des prédicteurs pour les trois types de modèles (IQH, MLG, ARS) montre que c'est le substrat qui a eu la plus grande influence sur la prédiction du modèle IQH (figure A.21). Les autres prédicteurs étaient cependant très influents également. En effet, tous les prédicteurs présentaient une pleine qualité sur une grande partie de la zone d'étude.

Au cours de l'examen du modèle préliminaire, il a été suggéré de créer des modèles distincts pour les juvéniles et les adultes. Toutefois, compte tenu des contraintes de temps, nous avons modélisé les juvéniles et les adultes ensemble et nous créerons des modèles distincts à une date ultérieure.

Le sébaste aux yeux jaunes est généralement présent depuis les îles Umnak et Unalaska, dans les îles Aléoutiennes, jusqu'à Ensenada, dans le nord de la Basse-Californie, et a été signalé entre 15 et 549 m de profondeur (Love *et al.*, 2002). Dans les eaux de la Colombie-Britannique, les observations du sébaste aux yeux jaunes dans les zones côtières varient de 31 à 250 m de profondeur (98 % de toutes les données de relevé de la base de données Ground Fish Bio – à l'exclusion des 1 % supérieur et inférieur). Compte tenu de ces aires de répartition indiquées, nous avons attribué une qualité augmentant de 0 à 1 pour la profondeur entre 15 et 31 m afin de permettre une transition d'une profondeur inadéquate à une profondeur pleinement convenable. La qualité reste à 1 de 31 à 250 m, puis diminue jusqu'à 0,1 à 300 m. Elle reste marginale à 0,1 jusqu'à 549 m de profondeur, après quoi elle tombe à 0. La diminution entre 250 et 300 m permet une transition entre une profondeur pleinement et marginalement appropriée, et correspond à la rétroaction sur l'examen préliminaire selon laquelle la diminution de la qualité de la profondeur est probablement plus que linéaire entre 250 et 549 m.

Les sébastes aux yeux jaunes préfèrent les habitats rocheux, mais peuvent occuper des substrats benthiques comme le sable et la vase (National Marine Fisheries Service, 2014). Nous attribuons donc une pleine qualité de 1 pour les substrats rocheux et mixtes, de 0,25 pour le sable (Haggarty) et une qualité marginale de 0,1 pour la vase.

En Colombie-Britannique, les sébastes aux yeux jaunes sont observés durant les relevés au chalut de fond à des températures variant de 4,56 à 16,8 °C et à des salinités (relevés au chalut de fond de 2013 à 2018 et relevé à la palangre de 2018) comprises entre 28,3 et 34 USP (Keppel⁶). Les plages de températures et de salinités indiquées pour le sébaste aux yeux jaunes dans le sud-est de l'Alaska (Johnson *et al.*, 2003), près de l'extrémité nord de leur aire de répartition, se situent dans les plages de températures observées en Colombie-Britannique. Nous n'avons pas pu trouver les plages de températures ou de salinités signalées près de la partie sud de leur aire de répartition. Étant donné le manque d'information sur les préférences et les valeurs optimales, nous attribuons une pleine qualité de 1 à l'ensemble des plages de températures et de salinités observées pour la Colombie-Britannique.

Le sébaste aux yeux jaunes préfère les habitats complexes comme la couche rocheuse fracturée et les complexes de blocs rocheux (p. ex. National Marine Fisheries Service, 2014; Yamanaka *et al.*, 2006). Pour exprimer cette complexité de l'habitat, nous avons d'abord créé un indice de qualité pour la rugosité, puis un indice pour la pente. Les indices ainsi obtenus étaient très restrictifs, produisant des modèles IQH avec très peu d'habitats convenables, car les données n'étaient pas à une échelle suffisamment fine pour refléter les processus agissant sur l'aire de répartition du sébaste aux yeux jaunes. Nous avons donc choisi de ne pas inclure les indices de rugosité et de pente dans le modèle.

Étoile ocrée

Pisaster ochraceus

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat et la profondeur dans le modèle IQH pour l'étoile ocrée (figure A.22). L'influence relative des prédicteurs montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle (figure A.23).

Les étoiles ocrées sont présentes du milieu de la zone intertidale jusqu'à 87 m de profondeur (Harbo, 2011), avec une plage optimale de -3 m (milieu de la zone intertidale) à 2 m (Harley⁷). Nous avons assigné une profondeur convenable entre -5 et 87 m, avec une augmentation linéaire de la qualité de 0 à 1 entre -5 et -3 m de profondeur, une qualité optimale de 1 entre -3 et 2 m de profondeur, puis une diminution linéaire jusqu'à une qualité marginale de 0,1 à 10 m de profondeur. La qualité augmente linéairement de 0 à 1 en haut de la plage de profondeurs pour refléter la transition entre les profondeurs inadéquates et les profondeurs convenables. De même, il y a une transition d'une profondeur entièrement convenable à une profondeur marginalement convenable juste en dessous de la plage de profondeurs optimale.

Ces étoiles de mer occupent le littoral rocheux et sont des prédateurs écologiquement importants dans une vaste gamme d'environnements, des lagunes abritées aux rivages les plus exposés aux vagues (Dayton, 1971; Levin et Paine, 1974). Pour les substrats, nous avons attribué une pleine qualité de 1 pour la roche, une qualité modérée de 0,5 pour les substrats mixtes et une qualité marginale de 0,1 pour le sable et de 0,05 pour les substrats boueux. Puisque les étoiles ocrées peuvent tolérer une vaste gamme d'énergies des vagues et des marées, la vitesse des courants de marée et l'exposition n'ont pas été incluses dans le modèle.

Les étoiles ocrées peuvent également vivre dans une plage étendue de salinités et de températures. Dans une étude sur l'effet des changements de salinité sur les populations vivant dans des milieux à salinité faible et élevée, Held et Harley (2009) ont utilisé des populations naturellement présentes à 20 USP (Vancouver, C.-B.) et 30 USP (Bamfield, C.-B.). Les auteurs ont constaté que les populations vivant dans les milieux peu salins ont pu survivre pendant de longues périodes à 15 USP, la salinité la plus faible analysée, et que les taux d'alimentation étaient les plus élevés pour chacune des populations à sa salinité *in situ*. Dans une étude sur les effets de la température, Gooding *et al.* (2009) ont constaté que les taux de croissance augmentaient linéairement avec des températures de 5 à 21 °C. Cependant, Sanford (2002) a également observé une plus forte croissance par gramme de tissu de moule consommé à 9 °C qu'à 12 °C, ce qui suggère que la réduction de la consommation dans des conditions plus froides est compensée par une réduction des coûts métaboliques. La limite létale pour la température corporelle est de 35,8 °C (Pincebourde *et al.*, 2008). Étant donné que les plages signalées se situent à l'intérieur de celles de la zone d'étude, la température et la salinité n'ont pas été incluses dans le modèle. Même si l'on rapporte que les taux de croissance augmentent linéairement avec des températures de 5 à 16 °C, on ne connaît pas exactement la plage optimale, et nous supposons que toutes ces températures conviennent à la survie.

⁷ Harley, C., Professeur, Département de zoologie de l'Université de la Colombie-Britannique et Institut des océans et des pêches. Avril 2019, communication personnelle.

Complexe de moules bleues

Mytilus edulis, *M. trossulus* et *M. galloprovincialis*

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur et la vitesse des courants de marée dans le modèle IQH pour le complexe de moules bleues (figure A.24). L'influence relative des prédicteurs montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle (figure A.25).

Dans le Pacifique, trois espèces ou sous-espèces sont considérées comme faisant partie du complexe *Mytilus edulis* : *M. edulis*, *M. galloprovincialis* et *M. trossulus*. Il n'est pas possible de distinguer ces espèces, capables de s'hybrider, à partir des seules caractéristiques de leur coquille (Harbo, 2011). Nous avons modélisé ces espèces ensemble.

Harbo (2011) indique que ces moules se trouvent depuis la zone intertidale jusqu'à une profondeur de 5 m, et Williams (1989) signale une profondeur maximale de 45 m, avec une colonisation dense entre des profondeurs de -3,7 m (haut de la zone mi-intertidale) et -1,5 m (bas de la zone intertidale). La profondeur a été assignée comme convenable dans la plage comprise entre -6 et 45 m, avec une augmentation linéaire de la qualité de 0 à 1 entre -6 et -5 m, une pleine qualité de 1 entre -5 et -1,5 m (bas de la zone intertidale), puis une diminution linéaire jusqu'à une qualité marginale de 0,1 par 5 m de profondeur. L'augmentation de la qualité au sommet de la plage de profondeurs et sa diminution entre -1,5 et 5 m de profondeur ont été incluses pour refléter les transitions entre les habitats entièrement et marginalement convenables. Au cours de l'examen préliminaire des prédicteurs, les examinateurs ont suggéré que la plage de profondeurs optimale pour les moules bleues se situerait à l'endroit où se trouve la petite balane, *Chthamalus dalli* (Norgard et Bigg⁸). Klinkenberg (2017) signale que *C. dalli* est présente en haut de la zone intertidale, et Morris *et al.* (1980) indiquent une occurrence du haut à la moitié de la zone intertidale. Pour refléter cette aire de répartition, nous avons étendu la plage de profondeurs optimale pour la moule bleue rapportée par Williams (1989; -3,7 à -1,5 m de profondeur) afin d'inclure le haut de la zone intertidale (jusqu'à -5 m de profondeur).

Les moules bleues forment des agrégations denses sur les surfaces dures (Harbo, 2011) et se trouvent sur les rochers, les piliers des jetées, les flotteurs, le gravier, la vase compacte et les organismes à coquille dure (Carlton, 2007; Morris *et al.*, 1980; Williams, 1989). Pour les substrats, nous avons attribué une pleine qualité de 1 pour la roche, une qualité modérée de 0,5 pour les substrats mixtes et une qualité marginale de 0,1 pour les substrats sablonneux et boueux.

Les moules peuvent tolérer une plage étendue de salinités et de températures, de 15 à 40 USP (Williams, 1989) et de 10 à 25 °C (Brenko et Calabrese, 1969), respectivement. La salinité et la température n'ont pas été incluses dans le modèle parce que les valeurs dans la zone d'étude se situaient à l'intérieur des limites indiquées.

Les moules bleues peuvent également tolérer une vaste gamme d'énergies des vagues et des marées, des estuaires abrités jusqu'aux côtes exposées (Williams, 1989). Toutefois, les examinateurs ont indiqué que certains endroits de la zone d'étude sont trop stagnants pour la survie (Norgard et Bigg⁸). Nous avons utilisé la vitesse du courant de marée pour exclure l'habitat stagnant, et nous avons assigné une qualité croissante de 0 à 1 entre 0,03 et 0,05 m·s⁻¹, avec une pleine qualité de 1 au-delà. L'exposition n'a pas été incluse dans le

⁸ Norgard, T., et Bigg, M. Section de l'écologie et analyse de l'espace marin, Sciences, Pêches et Océans Canada. Décembre 2018, communication personnelle.

modèle, car les moules bleues se trouvent dans des zones où l'énergie des vagues est très variable et nous avons pu exclure les régimes d'énergie les plus faibles à l'aide de la relation prédictive de la vitesse des courants de marée.

Palourde du Pacifique

Leukoma staminea

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur, la vitesse des courants de marée et l'exposition dans le modèle d'enveloppe IQH pour la palourde du Pacifique (figure A.26). L'influence relative des prédicteurs montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle (figure A.27).

La palourde du Pacifique habite des régions de sable stable, de vase tassée ou de mélanges d'argile et de gravier, et on la trouve parfois dans les sédiments graveleux parmi des roches (Williams, 1989). Nous avons attribué une pleine qualité de 1 pour les substrats de sable et de vase, une qualité moyenne de 0,5 pour les substrats mixtes et une qualité marginale de 0,1 pour la roche.

Harbo (2011) indique que cette palourde occupe le milieu de la zone intertidale jusqu'à une profondeur de 10,5 m, et Williams (1989) précise une plage de profondeurs allant de légèrement au-dessus du milieu de la zone intertidale jusqu'à 12 m. Pour la profondeur, nous avons utilisé une relation décroissante retardée, en supposant que ces palourdes se trouvent de préférence à l'extrémité peu profonde de leur plage, où les concentrations de nutriments et d'oxygène dissous sont plus élevées. Nous avons assigné une augmentation de la qualité de 0 à 1 entre -6 m (laisse des hautes eaux) et -3 m (moitié de la zone intertidale) de profondeur, puis une diminution retardée, jusqu'à une qualité de 0 à 12 m de profondeur.

La palourde du Pacifique adulte peut survivre à des températures comprises entre 0 et 25 °C (Chew et Ma, 1987) et à des salinités de 20 à 35 USP (Glude, 1978; Quayle et Bourne, 1972). La plage des valeurs moyennes de température et de salinité dans notre étude se situe entièrement à l'intérieur des plages de températures et de salinités établies, de sorte qu'aucun indice de qualité de la température ou de la salinité n'était requis dans le modèle.

La palourde du Pacifique se rencontre couramment sur les plages protégées des baies et estuaires où l'exposition est faible et où l'énergie des vagues et des courants est de faible à modérée (Williams, 1990). Pour les valeurs de l'exposition, nous avons créé une relation de seuil, avec une pleine qualité de 1 pour les valeurs de l'exposition basses jusqu'à 30 (en centaines de kilomètres), puis une diminution linéaire jusqu'à une qualité de 0 à la valeur modérée de 50. Cela a permis d'exclure les zones fortement exposées où les substrats meubles seraient probablement emportés par l'eau. De même, nous avons maintenu des valeurs de la vitesse des courants de marée faibles à modérées, avec une qualité de 1 jusqu'à 0,4 m·s⁻¹, puis une diminution linéaire jusqu'à une qualité de 0 à 0,81 m·s⁻¹.

Pennatule orangée

Ptilosarcus gurneyi

Compte tenu des connaissances écologiques décrites ci-après, nous avons inclus le substrat, la profondeur et la vitesse des courants de marée dans le modèle IQH pour la pennatule orangée (figure A.28). L'influence relative des prédicteurs montre que c'est la profondeur qui a eu la plus grande influence sur la prédiction du modèle (figure A.29).

La pennatule orangée est un anthozoaire colonial sessile que l'on trouve en agrégations denses dans des sédiments meubles (Lamb et Hanby, 2006). Compte tenu de cette préférence, nous attribuons une qualité de 1 pour les substrats de sable et de vase et de 0,5 pour les substrats mixtes. Dans les couches de prédicteurs du substrat de la zone d'étude du plateau, il y a plusieurs zones où l'on prévoit la présence de roches dans des zones connues pour être principalement constituées de sable ou de vase. Pour en tenir compte, nous attribuons une qualité de 0,25 pour le substrat rocheux afin que ces zones ne soient pas considérées comme totalement non convenables.

Harbo (2011) signale la présence de pennatules orangées depuis la zone intertidale jusqu'à 100 m de profondeur ou plus, et Lamb et Hanby (2006) en ont observé jusqu'à 135 m dans la zone infratidale. En combinant ces rapports, et en l'absence d'information sur les préférences en matière de profondeur, nous supposons qu'une qualité de 1 convient pour toute la plage de profondeurs signalée, du bas de la zone intertidale (-1 m) à 135 m.

Les pennatules orangées sont des suspensivores passifs et dépendent des courants ambiants qui poussent la nourriture sur un réseau de polypes sur des branches en forme de feuilles dans la partie supérieure de l'organisme. Dans une série d'expériences en laboratoire et sur le terrain près de Friday Harbour (État de Washington), Best (1988) a examiné la relation entre la vitesse ambiante (marée) et l'efficacité de l'alimentation chez les pennatules orangées. Il a estimé que l'efficacité de l'alimentation augmentait jusqu'à un maximum entre 8 et 20 $\text{cm}\cdot\text{s}^{-1}$, selon la taille de l'individu, et diminuait hors de cette plage. Compte tenu de ces estimations, nous attribuons une relation parabolique pour la vitesse avec une qualité optimale de 1 à 14 $\text{cm}\cdot\text{s}^{-1}$.

Nous n'incluons pas les prédicteurs de salinité ou de température, car nous manquons d'information sur les seuils et les préférences concernant ces deux variables environnementales pour cette espèce.

RÉFÉRENCES ANNEXE B

- Best, B.A. 1988. Passive suspension feeding in a sea pen: Effects of ambient flow on volume flow rate and filtering efficiency. *Biological Bulletin* **175**: 332-342.
- Breen, P., and Adkins, B. 1979. A survey of abalone populations on the east coast of the Queen Charlotte Islands, August 1978. *Fish. Mar. Serv. Manuscr. Rep* **1490**: 125.
- Breen, P.A. 1986. Management of the British Columbia fishery for Northern Abalone (*Haliotis kamtschatkana*). *Can. Spec. Pub. Fish. Aquat. Sci.* **92**: 300-312.
- Brenko, M.H., and Calabrese, A. 1969. The combined effects of salinity and temperature on larvae of the mussel *Mytilus edulis*. *Mar. Biol.* **4**(3): 224-226.
- Campbell, A., Harbo, R.M., and Hand, C.M. 1998. Harvesting and distribution of Pacific geoduck clams, *Panopea abrupta*, in British Columbia. In *Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management*. Edited by G.S. Jamieson and A. Campbell. NRC Research Press, Ottawa, ON. pp. 349-358.
- Carlton, J.T. 2007. *The Light and Smith Manual: Intertidal Invertebrates from Central California to Oregon*. University of California Press, Berkeley, California.
- Chew, K.K., and Ma, A.P. 1987. Species profiles: Life histories and environmental requirements of coastal fishes and invertebrates (Pacific Northwest) - common littleneck clam. U.S. Fish and Wildlife Service. 22 pp.
- Cleaver, F.C. 1949. Preliminary results of the coastal crab (*Cancer magister*) investigation. Department of Fisheries, Washington State. 47–92 pp.
- COSEPAC. 2009. Évaluation et Rapport de situation du COSEPAC sur l'Ormeau nordique *Haliotis kamtschatkana* au Canada – Mise à jour. Comité sur la situation des espèces en péril au Canada. Ottawa. vii + 51 p.
- Curtis, D.L., and McGaw, I.J. 2008. A year in the life of a Dungeness crab: methodology for determining microhabitat conditions experienced by large decapod crustaceans in estuaries. *J. Zool.* **274**: 375–385.
- Curtis, D.L., and McGaw, I.J. 2012. Salinity and thermal preference of Dungeness crabs in the lab and in the field: Effects of food availability and starvation. *Journal of Experimental Marine Biology and Ecology* **413**: 113-120.
- Dayton, P.K. 1971. Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. *Ecological Monographs* **41**: 351-389.
- DFO. 2017. Pacific Region Integrated Fisheries Management Plan: Crab by Trap - January 1, 2017 to March 31, 2018.
- DFO. 2018. 2018/2019 Red Sea Urchin Integrated Fisheries Management Plan.
- Druehl, L.D., and Clarkston, B. 2016. *Pacific Seaweeds: A guide to the common seaweeds of the West Coast*. Harbour Publishing, Madeira Park, BC.
- Emmett, B., and Jamieson, G.S. 1988. An experimental transplant of northern abalone, *Haliotis kamtschatkana*, in Barkley Sound, British Columbia. *Fishery Bulletin, U.S.* **87**: 95-104.
- Engelhardt, F.R., and Dehnel, P.A. 1973. Ionic regulation in Pacific edible crab, *Cancer magister* (Dana). *Canadian Journal of Zoology* **51**: 735–743.

-
- Glude, J.B. 1978. The clam genera *Mercenaria*, *Saxidomus*, *Protothaca*, *Tapes*, *Mya*, *Panope*, and *Spisula*: a Literature review and analysis of the use of thermal effluent in the culture of clams. Aquaculture Consultant Report. 74 pp.
- Gooding, R.A., Harley, C.D.G., and Tang, E. 2009. Elevated water temperature and carbon dioxide concentration increase the growth of a keystone echinoderm. *PNAS* **106**(23): 9316-9321.
- Goodwin, C.L. 1973. Effects of salinity and temperature on embryos of the geoduck clam (*Panopea generosa*). Proceedings of the National Shellfish Association. pp. 93-95.
- Goodwin, C.L., and Pease, B. 1989. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Pacific Northwest) - Pacific geoduck clam. U.S. Fish and Wildlife Service 11.120.
- Harbo, R.M. 2011. *Whelks to Whales: Coastal Marine Life of the Pacific Northwest*. Harbour Publishing, Madiera Park, B.C.
- Hawkes, M.W., Tanner, C.E., and Lebednik, P.A. 1979. The benthic marine algae of northern British Columbia. *Syesis* **11**: 81-115.
- Held, M.B.E., and Harley, D.G.H. 2009. Responses to low salinity by the sea star *Pisaster ochraceus* from high- and low-salinity populations. *Invertebr. Biol.* **128**(4): 381–390.
- Johnson, S.W., Murphy, M.L., and Csepp, D.J. 2003. Distribution, habitat, and behavior of rockfishes, *Sebastes* spp., in nearshore waters of southeastern Alaska: observations from a remotely operated vehicle. *Environ. Biol. Fishes* **66**: 259–270.
- Klinkenberg, B. 2017. *E-Fauna BC: Electronic Atlas of the Fauna of British Columbia*. Available from efauna.bc.ca.
- Kondzela, C.M., and Shirley, T.C. 1993. Survival, feeding, and growth of juvenile Dungeness crabs from southeastern Alaska reared at different temperatures. *J. Crust. Biol.* **13**: 25–35.
- Lamb, A., and Hanby, B. 2006. *Marine Life of the Pacific Northwest: A Photographic Encyclopedia of Invertebrates, Seaweeds and Selected Fishes*. Harbour Publishing, Madiera Park, B.C.
- Lessard, J., and Campbell, A. 2007. Describing northern abalone, *Haliotis kamtschatkana*, habitat: focusing rebuilding efforts in British Columbia, Canada. *J. Shellfish Res.* **26**(3): 677-686.
- Levin, S.A., and Paine, R.T. 1974. Disturbance, patch formation, and community structure. *PNAS* **71**: 2744-2747.
- Love, M.S., Yoklavich, M., and Thorsteinson, L. 2002. *The Rockfishes of the Northeast Pacific*. University of California Press, Berkeley, California.
- McGaw, I.J., Reiber, C.L., and Guadagnoli, J.A. 1999. Behavioral physiology of four crab species in low salinity. *Biological Bulletin* **196**: 163–176.
- Mondragon, J., and Mondragon, J. 2003. Seaweeds of the Pacific Coast: Common Marine Algae from Alaska to Baja California. *Sea Challengers*, Monterey, CA. pp. 97.
- Morris, R.H., Abbott, D.L., and Haderlie, E.C. 1980. *Intertidal invertebrates of California*. Stanford University Press, Stanford, CA.
- MPO. 2007. Programme de rétablissement de l'haliotide pie (*Haliotis kamtschatkana*) au Canada. Série de Programmes de rétablissement de la Loi sur les espèces en péril. Pêches et Océans Canada, Vancouver. vi + 31 pp.

-
- National Marine Fisheries Service. 2014. Designation of Critical Habitat for the Distinct Population Segments of Yelloweye Rockfish, Canary Rockfish, and Bocaccio. NMFS, Protected Resources Division, West Coast Region, Seattle, WA. 77 pp.
- Paul, A.J., and Paul, J.M. 1998. Respiration rate and thermal tolerances of pinto abalone (*Haliotis kamtschatkana*). *J. Shellfish Res.* **17**: 743-745.
- Pauley, G.B., Armstrong, D.A., Van Citter, C., and Thomas, G.L. 1989. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Pacific Southwest)—Dungeness crab. U.S. Fish and Wildlife Service. Biological Report 11.121.
- Pêches et Océans Canada. 2007. Programme de rétablissement de l'haliotide pie (*Haliotis kamtschatkana*) au Canada. Série de Programmes de rétablissement de la Loi sur les espèces en péril. Pêches et Océans Canada, Vancouver. vi + 31 pp.
- Phillips, R.C., and Lewis, R.L. 1983. Influence of environmental gradients on leaf widths and transplant success in North American seagrasses. *Mar. Technol. Soc. J.* **17**: 59-68.
- Pincebourde, S., Sanford, E., and Helmuth, B. 2008. Body temperature during low tide alters the feeding performance of a top intertidal predator. *Limnol. Oceanogr.* **23**: 1562-1573.
- Quayle, D.B., and Bourne, N. 1972. The clam fishery of British Columbia. Fisheries Research Board of Canada Bulletin **179**: 70.
- Richards, L.J. 1986. Depth and habitat distributions of three species of rockfish (*Sebastes*) in British Columbia: observations from the submersible PISCES IV. *Environ. Biol. Fishes* **17**(1): 13-21.
- Richards, L.J., and Hand, C.M. 1987. 1987 research catch and effort data on nearshore reef-fishes in British Columbia statistical areas 12 and 13. *Can. Man. Rep. Fish. Aquat. Sci.* **1958**:59.
- Sanford, E. 2002. The feeding, growth, and energetics of two rocky intertidal predators (*Pisaster ochraceus* and *Nucella canaliculata*) under water temperatures simulating episodic upwelling. *J. Exp. Mar. Biol. Ecol.* **273**: 199–218.
- Sloan, N.A., and Breen, P.A. 1988. Northern Abalone, *Haliotis kamtschatkana*, in British Columbia: Fisheries and Synopsis of Life History Information. *Can. Spec. Pub. Fish. Aquat. Sci.* **103**.
- Stone, R. P., and C. E. O'Clair (2001), Seasonal movements and distribution of Dungeness crabs, *Cancer magister*, in a glacial southeastern Alaska estuary, *Mar. Ecol. Prog. Ser.*, **214**: 167-176.
- Tomascik, T., and Holmes, H. 2003. Distribution and abundance of *Haliotis kamtschatkana* in relation to habitat, competitors and predators in the Broken Group Islands, Pacific Rim National Park Reserve of Canada. *J. Shellfish Res.* **22**: 831-838.
- Vandermeulen, H. 2005. Assessing marine habitat sensitivity: a case study with eelgrass (*Zostera marina* L.) and kelps (*Laminaria*, *Macrocystis*). DFO Can. Sci. Advis. Sec. Res. Doc. 2005/032 57 pp.
- Williams, G.L. 1989. Coastal/estuarine fish habitat description & assessment manual Part I: Species/Habitat Outlines. Williams & Associates Ltd., Coquitlam, BC. 139 pp.
- Williams, G.L. 1990. Coastal/estuarine fish habitat description & assessment manual Part II: Habitat Description Procedures. Williams & Associates Ltd., Coquitlam, BC. 61 pp.

Yamanaka, K.L., Lacko, L.C., Miller-Saunders, K., Grandin, C., Lochead, J.K., Martin, J.C., Olsen, N., and Wallace, S.S. 2006. A review of quillback rockfish, *Sebastes maliger*, along the Pacific coast of Canada: biology, distribution and abundance trends. DFO Can. Sci. Advis. Sec. Res. Doc. 2006/077. 68 pp.