



Modélisation et interpolation spatiale 3D pour l'étude de l'écosystème pélagique marin

Mémoire

Per Jonas Sahlin

Maîtrise en sciences géomatiques
Maître ès Sciences (M. Sc.)

Québec, Canada

© Jonas Sahlin, 2014

Résumé

En raison de la nature dynamique et volumétrique de l'écosystème marin pélagique, sa modélisation spatiale constitue un défi important. La représentation conventionnelle des phénomènes de ce milieu s'effectue par des coupes statiques verticales ou horizontales dans un environnement bidimensionnel (2D). Cependant, comme le démontre ce mémoire, l'étude de l'écosystème marin peut être grandement améliorée grâce à la modélisation spatiale tridimensionnelle (3D). L'apport principal de cette étude est d'avoir démontré le potentiel et la pertinence de la modélisation spatiale 3D pour l'environnement pélagique marin. Cette étude confirme que les outils émergents de visualisation scientifique dans le domaine de la modélisation géologique peuvent servir à améliorer l'étude de cet écosystème. Elle permet également de combler une lacune importante identifiée dans la littérature scientifique en examinant la performance des méthodes d'interpolation spatiale 3D. L'interpolation spatiale est une étape essentielle de la modélisation spatiale 3D des phénomènes continus (p.ex. salinité, température, etc.), mais aucune étude n'avait encore évalué son efficacité pour l'environnement pélagique marin. Il s'agit donc d'un pas important vers le développement d'un système d'information géographique (SIG) marin 3D complet. Les avantages de migrer vers une modélisation spatiale 3D sont discutés dans le contexte de la campagne océanographique ArcticNet-Malina, réalisée dans la mer de Beaufort (Arctique canadien) en 2009. Des représentations spatiales 3D basées sur une stratégie d'interpolation 3D *robuste* et *optimale* de cinq variables pélagiques marines (température, concentration en chlorophylle a, coefficient d'atténuation particulaire, distribution des eaux de l'halocline supérieure et flux vertical de carbone organique particulaire) sont présentées et leurs valeurs écologiques sont discutées.

Abstract

Spatial modeling of the marine pelagic ecosystem is challenging due to its dynamic and volumetric nature. Consequently, conventional oceanographic spatial analysis of this environment is in a 2D environment, limited to static cutting planes in horizontal and vertical sections to present various phenomena. However, the study of the marine pelagic ecosystem can benefit from 3D spatial modeling. The main contribution of this study is to show that recent advances in 3D spatial modeling tools developed primarily for geological modeling can be exploited to extend the usual interpretation of marine pelagic phenomena from a 2D to a 3D environment. This study also fills a major gap identified in the literature by examining the performance of 3D spatial interpolation methods. Such interpolation is an essential step in 3D spatial modeling of continuous phenomena (eg, salinity, temperature, etc.), but no study has yet evaluated its performance for the marine pelagic environment. Accordingly, this study constitutes an important step towards the development of a complete 3D marine GIS. The benefits of migrating to a 3D spatial modeling of the marine environment are discussed in the context of the oceanographic campaign ArcticNet-Malina, conducted in the Beaufort Sea (Canadian Arctic) in 2009. 3D spatial representations based on a robust and optimal 3D interpolation strategy for five pelagic variables (temperature, chlorophyll *a*, particulate attenuation coefficient, distribution of upper halocline water mass and vertical flux of particulate organic carbon) of the ArcticNet-Malina campaign are presented and their ecological values are discussed.

Table des matières

RÉSUMÉ.....	III
ABSTRACT.....	V
TABLE DES MATIÈRES.....	VII
LISTE DES TABLEAUX.....	IX
LISTE DES FIGURES.....	XI
REMERCIEMENTS.....	XV
AVANT-PROPOS.....	XVII
1 CHAPITRE 1 – INTRODUCTION.....	1
1.1 Mise en contexte	1
1.2 Étude de cas - Mer de Beaufort, Arctique canadien.....	3
1.3 Problématique	5
1.3.1 Problématique générale.....	5
1.3.2 Problématique spécifique.....	6
1.4 Objectifs	7
1.5 Méthodologie	9
1.6 Structure du mémoire.....	11
2 CHAPITRE 2 - REVUE DES CONCEPTS.....	13
2.1 L'écosystème pélagique marin	13
2.1.1 Délimitation.....	13
2.1.2 Forçages océaniques et variabilité spatio-temporelle.....	13
2.1.3 Caractéristiques particulières d'un point de vue de la géomatique	15
2.1.4 Méthodes d'acquisition de données.....	17
2.2 Modélisation spatiale 3D de l'écosystème pélagique.....	18
2.2.1 Conceptualisation de l'espace et modélisation spatiale.....	18
2.2.2 Modélisation spatiale 3D des variables continus.....	21
2.2.3 Outils d'analyse de modèles 3D de phénomènes continus.....	24
2.2.4 Interpolation spatiale.....	26
2.2.5 Visualisation de volumes.....	38
3 CHAPITRE 3 - 3D GEOSPATIAL MODELLING AND VISUALIZATION FOR MARINE ENVIRONMENT: STUDY OF THE MARINE PELAGIC ECOSYSTEM OF THE SOUTH-EASTERN BEAUFORT SEA, CANADIAN ARCTIC.....	41
3.1 Résumé.....	41
3.2 Abstract	42
3.3 Introduction	43
3.4 3D spatial representation of the pelagic ecosystem.....	44
3.5 3D geospatial solution to representation of water masses distribution: Malina case study.....	46

3.6	Discussion and conclusions	47
3.7	Acknowledgements	47
4	CHAPITRE 4 - ASSESSMENT OF 3D SPATIAL INTERPOLATION METHODS FOR STUDY OF THE MARINE PELAGIC ENVIRONMENT	49
4.1	Résumé	49
4.2	Abstract	50
4.3	Introduction	51
4.4	Background	53
4.4.1	Inverse Distance Weighting	54
4.4.2	Ordinary Kriging and 3D variogram	55
4.4.3	Accounting for the dynamic nature of the marine pelagic environment	56
4.5	Methods	57
4.5.1	Study area and data preparation	57
4.5.2	Computational details	59
4.5.3	Method performance comparison criteria	62
4.6	Results	64
4.6.1	Variogram modeling results	64
4.6.2	CV error results	66
4.6.3	Visual examination of 3D spatial representations	69
4.7	Discussion	70
4.7.1	3D variogram and spatial autocorrelation	70
4.7.2	Method performance	71
4.7.3	Clustered sampling and temporal variation influences	74
4.7.4	Final volumetric models and ecological importance	75
4.8	Conclusions	79
4.9	Acknowledgements	81
5	CHAPITRE 5 - CONCLUSION GÉNÉRALE	83
5.1	Contribution de la recherche	83
5.2	Atteinte des objectifs	83
5.3	Perspectives et futurs travaux	85
5.3.1	Développement d'un SIG marin 3D	85
5.3.2	Modélisation et interpolation spatiale 3D pour l'écosystème pélagique marin	86
	BIBLIOGRAPHIE	89
	ANNEXE A. VARIOGRAMMES 3D	97
	ANNEXE B. RÉSULTATS DE LA VALIDATION CROISÉE (CV) DES STRATÉGIES D'INTERPOLATION	103
	ANNEXE C. ÉVALUATION QUALITATIVE DES STRATÉGIES D'INTERPOLATION 3D ..	105

Liste des tableaux

Tableau 1.	Critères d'évaluation de logiciels de modélisation spatiale 3D de phénomènes continus.....	25
Tableau 2.	Review of five geospatial modelling tools from commercial and academic GIS as well as from geomodelling.....	45
Tableau 3.	Basic statistics for variables and data sets used in research.....	59
Tableau 4.	Variation of input parameters for the assessment of robustness of interpolation strategies. n refers to the number of sampling points used for the weighted estimation, <i>horizontal ranges</i> refer to distances of spatial autocorrelation in the horizontal plane derived from variogram modeling and ρ corresponds to the IDW-power.	62
Tableau 5.	Variogram modeling results. Spherical model used for all variograms.	66
Tableau 6.	Performance des méthodes d'interpolation 3D sur la base de métriques d'erreur MSE et MAE. n représente le voisinage de l'interpolation, les valeurs de MSE et MAE correspondent au métrique d'erreur moyenne d'une interpolation par rapport aux valeurs mesurées (voir texte section 4.5.3.).	103

Liste des figures

- Figure 1. Représentation habituelle de phénomènes océanographiques. (a) Section horizontale représentant les variations spatiales de la concentration en chlorophylle a (chl a) (profondeur de 3 m) dans l'est de Great Australian Bight (Australie) (tiré de van Ruth, Ganf, et Ward 2010). (b) Section verticale de chl a dans la mer de Beaufort (tiré de (Martin et al. 2010). L'axe horizontal de la Figure 1b correspond à la distance de la côte (km), alors que les marqueurs sur cet axe correspondent aux stations d'échantillonnage..... 3
- Figure 2. Région d'étude de la campagne océanographique ArcticNet-Malina (2009)..... 4
- Figure 3. Méthode pour la réalisation du projet de recherche..... 10
- Figure 4. Exemple de stratification dans l'écosystème pélagique marin. Coupe verticale de la masse d'eau de l'halocline supérieure (UHW) dans la mer de Beaufort. Le dégradé de couleurs représente la fraction de l'écosystème pélagique constituée de l'UHW. La zone grise représente le plateau continental. La figure est tirée de Lansard, Mucci et al. (2012)..... 14
- Figure 5. Étapes pour la création du modèle spatial de phénomènes géographiques (adapté de Burrough et Macdonnell 1998; Hashemi Beni 2009)..... 19
- Figure 6. Modèles de données raster 3D avec voxels de forme régulière (*gauche*) et irrégulière (*droite*)...22
- Figure 7. Triangulation de Delaunay (TD) et diagramme de Voronoï (DV). Le TD (*a*) et DV (*b*) pour un même ensemble de points en 2D. Dans *a*, les points (non illustrés) sont localisés aux vertex des triangles.....24
- Figure 8. Le variogramme. Exemple de variogramme avec 3 diagrammes de dispersion d'*h* pour trois intervalles d'*h* (*lags*) distincts. Les points correspondent au variogramme *expérimental*, alors que la ligne entière représente l'ajustement du variogramme (modèle sphérique). La variable (c_p) a été transformée en une distribution gaussienne normale.....31
- Figure 9. Variogramme directionnel, cas 2D. (a) Nomenclature du variogramme directionnel où lag step correspond à un intervalle d'*h*; Band width à la largeur du cône de tolérance. (b) Ellipse illustrant les portées de dépendance spatiale déduites d'un variogramme directionnel 2D.33
- Figure 10. Stratification de l'espace de modélisation pour l'estimation du variogramme 1D+2D. (Bergacademie XXXX).....34
- Figure 11. Variogramme 3D. Variogramme modélisé dans quatre intervalles d'azimuts (azimut central : 0 °, 45 °, 90 °, 135°) et trois intervalles d'angles de rotation verticale (0 °, 45 °, 85 °). Modèle du

variogramme = sphérique. La variable (c_p) a été transformée en une distribution gaussien normalisée
L'axe Y correspond au γ . L'axe X représente la fraction de l'espace de modélisation (*grid*) utilisée pour la
modélisation du variogramme.....35

Figure 12. Visualisation d'un volume par image 2D (température de l'eau, °C).....39

Figure 13. Exemple d'isosurfaces pour trois températures de l'eau différents. Températures : 7, 8 et 9°C
pour (a), (b) et (c). La Figure a été tirée de Ledoux et Gold 2008.....40

Figure 14. Combinaison des méthodes de visualisation : coupes verticales 2D, coupes 2.5D et une image
satellite. Le plan 2D horizontal représente le carbone organique particulaire (COP) de surface, généré à
partir d'un traitement d'images satellites. Les coupes verticales et la surface 2,5D collée sur le fond
bathymétrique sont issues d'un modèle volumétrique du flux de COP dans l'ensemble de la région
d'étude (*Ordinary Kriging*). Cette représentation spatiale a été effectuée avec Entervol for ArcGIS
(société CTech). Voir également Forest et al. 2013, *matériel complémentaire* pour un modèle interactif
de cette image.....40

Figure 15. Location of sampling stations for the Malina cruise in the Beaufort Sea, Canadian Arctic.....44

Figure 16. Kriging variance of spatial 3D model for pacific Upper Halocline Water in the Beaufort Sea.
Black dots indicate sampling points.....46

Figure 17. Geospatial representation of Upper Halocline Water on the Mackenzie shelf. Model constructed
with Paradigm Gocad). Scale indicate fraction (%) of water constituted of UHW. Voxels containing more
than 60 % of UHW are coloured-filled. Black dots indicate sampling points. Vertical exaggeration in
figure is 75 times that of reality.....47

Figure 18. Location of sampling stations used for the spatial modeling of pelagic variables as visited during
the ArcticNet-Malina oceanographic campaign, Beaufort Sea (Canadian Arctic) in 2009.....58

Figure 19. Grid designs assessed. Principal grid design used in study (left) and alternative grid design
(right).60

Figure 20. Anisotropic 3D variogram for c_p with fitted spherical model. Y-axes correspond to semi-
variogram variance (γ), while X-axes correspond to distance in grid units. The twelve interrelated
variograms are used to identify directions (azimuth, dip) of minimum and maximum spatial continuity.
Theoretical ranges of spatial autocorrelation is obtained from the X-axes where theoretical variogram
model (solid line) reaches a sill value of 1 ($\gamma = 1$).61

- Figure 21. Performance of interpolation strategies (n). MAE (X-axes) and MSE (Y-axes) correspond to mean absolute error and mean square error obtained through CV assessment. IDW with restricted vertical neighborhood (IDW_{uvw}) is illustrated for UHW. n represents the number of neighboring points used in estimations and p refers to the IDW power. Assessment of robustness of OK interpolations was made by an *underestimation* or *exaggeration* of horizontal variogram ranges by a factor of 1.5.....68
- Figure 22. Qualitative comparison of 3D geospatial representations of sea temperature in Beaufort Sea derived from four interpolation method strategies. Range of depth in spatial model is 0 to 1700 m, vertical exaggeration = 50. (a) Natural Neighbor interpolation (ontrol), (b) IDW, no vertical exaggeration of coordinates, $p = 5, n = 8$, (c) OK, $n = 32$. (d) IDW in uvw space (exaggeration of vertical coordinates,) $p = 5, n = 8$69
- Figure 23. Relative MAE of interpolation strategies with an optimal searching neighborhood (n). Y-axes represents the *relative MAE* ($MAE / \text{arithmetic mean of variable}$) and vertical bar represents the approximate average of both OK and IDW performance in a comparative study of 2D CV-assessments conducted by Li and Heap (2011).....72
- Figure 24. Volumetric model of temperature, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. 3D interpolation strategy: OK, $n = 32$. Depth represented of spatial model is 0 to 1700 m. Volume represents grid cells where temperature is superior to $1^{\circ}C$. Vertical exaggeration = 50 in figure... ..76
- Figure 25. Volumetric model of UHW, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. Fraction of water mass constituted of UHW (%) for the upper 500 m of water column. Interpolation: OK, $n=32$. Volume represents voxels where UHW fraction is $>50\%$. Black dots indicate sampling points. Areas with no-data values are colored transparent mauve. Vertical exaggeration below sea-surface = 200.77
- Figure 26. Volumetric model of chl a et c_p , during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. Spatial models represent 200 upper meters of water-column. Black dots indicate sampling points. Vertical exaggeration = 200. (a) Chl a , interpolation strategy = OK, $n = 32$. Volume for voxels illustrated when chl $a > 1 \text{ mg m}^{-3}$ (b) C_p , interpolation strategy = IDW_{xyz} , $p = 5, n = 48$. Volume for voxels illustrated when $c_p > 1 \text{ m}^{-1}$78
- Figure 27. Volumetric model of vertical POC-fluxes, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. 3D interpolation strategy: IDW_{xyz} , $p = 2, n = 48$. Upper 1700 m of water-column illustrated. Horizontal cuts are situated at 600 m depth in offshore areas and at 110 in shelf area (or at bottom for depths $<110 \text{ m}$). Black dots indicate sampling points. Vertical exaggeration = 50.....79

Figure 28.	Variogramme pour température (sous-ensemble).....	97
Figure 29.	Variogramme pour température (jeu de données complet).....	98
Figure 30.	Variogramme pour UHW (<i>grid</i> parallèle).....	99
Figure 31.	Variogramme pour UHW (<i>grid</i> proportionnel).....	100
Figure 32.	Variogramme pour chl <i>a</i>	101
Figure 33.	Variogramme pour flux vertical de COP	102
Figure 34.	Représentations spatiales 3D de la distribution de UHW (%). Les 500 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200.....	106
Figure 35.	Représentations spatiales 3D de chl <i>a</i> (mg m ⁻³). Les 200 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200.....	107
Figure 36.	Représentations spatiales 3D de c_p (m ⁻¹). Les 200 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200.....	108
Figure 37.	Représentations spatiales 3D du flux vertical de COP (mg C m ⁻² j ⁻²). Les 1700 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 50	109

Remerciements

En premier lieu, j'aimerais remercier Mir Abolfazi Mostafavi, mon directeur de thèse. En m'acceptant comme étudiant, il a aussi consenti à des contraintes particulières d'éloignement et il a dû avoir confiance en ma détermination et mon autonomie. Je le remercie pour cette confiance, ses encouragements et sa rigueur scientifique. J'adresse ensuite mes remerciements à Marcel Babin, codirecteur de ce mémoire, qui m'a accueilli dans l'environnement stimulant de l'équipe de sa chaire de recherche. J'aimerais aussi remercier plus particulièrement Alexandre Forest de l'équipe de Takuvik, mon premier interlocuteur tout au long du projet. Il a cru en ce projet depuis le départ, et sans ses encouragements et nos discussions stimulantes, je n'aurais pas mené cette recherche à terme. Merci également à Alfonso Condal pour avoir accepté d'agir à titre d'examineur externe de ce mémoire, ainsi qu'à la direction de GENIVAR et le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) pour leur support financier.

J'aimerais également remercier mes amis qui m'ont supporté pendant ces années d'étude et plus particulièrement : Catherine pour son écoute et ses précieuses corrections linguistiques, Hugo pour son regard rationnel, ses accords de blues en mineur mélodique et son partage de la passion pour le whiskey, ainsi que Nico pour les discussions scientifiques entremêlées des séances intenses et libératrices de percussions africaines.

Merci à ma famille – parents et sœurs – de l'autre côté de l'Atlantique pour leur confiance en moi et leur compréhension vu toutes ces conversations que j'ai coupées court en raison du faible nombre d'heures comprises dans un jour de notre système planétaire. Merci à mes trois petits soleils qui grandissent si vite et si bien, Maya, Yasmine et Lilou. Merci pour vos sourires et chaleureux câlins qui donnent un sens à la vie, qui m'apportent plus d'énergie que trois espressos doubles, plus de bonheur que n'importe quel mémoire achevé. Merci pour votre patience lorsque la mienne était épuisée, votre amour et votre envie de m'amener sur vos aventures imaginaires.

Mais surtout et avant tout, j'adresse mes remerciements à mon livskamrat et épouse adorable, Stéphanie, qui m'a toujours encouragé dans tous mes projets personnels et professionnels. C'est grâce à ton support que j'ai pu me discipliner pour achever cette aventure à travers les responsabilités professionnelles et familiales. C'est aussi toi qui m'as supporté, encouragé et enduré lors des moments d'hésitation et de fatigue après de longues journées, nuits et fins de semaine d'étude. Merci également pour ta généreuse relecture de mes textes. Je suis pleinement reconnaissant de l'ensemble de ton temps, tes efforts, ta patience, ton immense générosité, ta belle énergie et ton amour. Encore une fois Merci.

Avant-Propos

Ce mémoire avec insertion d'article est composé de six chapitres. Les chapitres 1, 2 et 5 constituent respectivement l'introduction, la revue des concepts ainsi que la conclusion du mémoire et sont composés en français. Les chapitres 3 et 4 constituent des articles scientifiques et ont été rédigés en anglais. Le chapitre 3 est un résumé long publié avec comité de lecture. Cet article a également été présenté lors de la conférence internationale du 7th 3DGeoInfo tenue à Québec au mois de mai 2012. Sa référence est la suivante :

Sahlin J., Mostafavi M. A., Forest A., Babin M. et Lansard B. (2012). 3D geospatial modelling and visualization for marine environment: Study of the marine pelagic ecosystem of the south-eastern Beaufort Sea, Canadian Arctic. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, Québec, ISPRS Archives, Copernicus Publications. Volume XXXVIII-4/C26, 21-24. DOI : 10.5194/isprsarchives-XXXVIII-4-C26-21-2012

Le chapitre 4 constitue le cœur du mémoire et constitue un article qui a été accepté pour publication dans le journal *Marine Geodesy* avec comité de lecture. Sa référence temporaire est la suivante :

Sahlin J., Mostafavi M. A., Forest A., Babin M. (2014). *Accepté pour publication*. Assessment of 3D spatial interpolation methods for study of the marine pelagic environment. *Marine Geodesy*, special issue on Marine and Coastal Geographical Information Systems (GIS) in Vol. 37, No. 2, 2014. XX p.

Tout comme ce mémoire, le travail de ces deux articles a été fait entièrement par Jonas Sahlin; je suis par conséquent l'auteur principal des ouvrages, réalisés sous la direction de Mir Abolfazl Mostafavi et la codirection de Marcel Babin. Les jeux de données utilisés dans le cadre de cette recherche viennent des campagnes océanographiques effectuées dans la mer de Beaufort en 2009, principalement dans le cadre du projet ArcticNet-Malina. Ces données ont été généreusement partagées par l'équipe de recherche Takuvik *Joint International Laboratory*, par l'entremise d'Alexandre Forest, qui a aussi produit des données sur le flux vertical de carbone organique. Les données sur la distribution des eaux de l'halocline supérieure proviennent de Bruno Lansard du *Laboratoire d'Études en Géophysique et Océanographie spatiales*, CNRS-UPS-CNRS (France).

CHAPITRE 1 – Introduction

1.1 Mise en contexte

Les océans couvrent 71 % de la surface de la Terre et leur profondeur moyenne est d'environ 4000 mètres; le volume de l'environnement marin pélagique (colonne d'eau) correspond à 99 % de la biosphère (Angel 1993). Le phytoplancton marin est responsable de la moitié de la production primaire¹ sur la Terre (Boyce, Lewis et al. 2010) et on y trouve plus de la moitié des poissons consommés par les humains (Pauly et al. 2002). L'importance économique et écologique des eaux marines pélagiques n'est donc pas à démontrer. Les enjeux environnementaux et socio-économiques auxquels ces espaces sont associés, et en particulier les zones côtières, sont également de plus en plus complexes, alors que près de la moitié de la population mondiale y vit et que les activités humaines s'y intensifient (Creel 2003). À ces enjeux, on peut ajouter l'effet des changements climatiques sur les écosystèmes marins et l'effet réciproque des masses d'eau sur l'atmosphère : il est généralement admis que notre compréhension du rôle joué par l'océan comme régulateur du climat de la planète est encore à ses balbutiements (Valdés, Fonseca, et Tedesco 2010). La gestion de ces milieux complexes présente donc un défi de taille et repose « *sur la prestation de conseils scientifiques de grande qualité, actuels et pertinents* » (Pêches et Océans Canada 2008).

Au niveau de la représentation spatiale, l'environnement pélagique marin est sensiblement différent de l'écosystème terrestre. Alors que ce dernier est souvent considéré comme un espace vide, largement 2D, rempli d'objets (maisons, arbres, etc.), l'écosystème pélagique marin est un vaste continuum de variables biotiques et abiotiques dans un environnement dynamique 3D. Les variations des caractéristiques physico-chimiques (p. ex. : température, salinité, phénomènes de remontée d'eau, fronts thermiques, masses d'eau, etc.) conditionnent l'organisation spatiale des variables biologiques (p. ex. la production de phytoplancton), des habitats pélagiques (Game et al. 2009; Kracker 1999) et de la présence d'organismes marins. L'étude de la distribution verticale et horizontale de ces variables physico-chimiques constitue la base de toute recherche océanographique.

L'acquisition de données sur l'environnement pélagique marin s'effectue traditionnellement le long de radiales, souvent perpendiculaires aux plateaux continentaux lorsqu'il s'agit d'environnements côtiers. Principalement en raison de la logistique et des coûts associés à ce genre de campagnes d'échantillonnage, le nombre de stations d'échantillonnage est limité tant dans le temps que dans l'espace. Ces limites imposent une représentation spatiale des données souvent déficiente, ce qui est aggravé par la nature dynamique et

¹ Quantité totale de matière organique fixée dans la photosynthèse chlorophyllienne (OQLF 2013).

trimensionnelle (3D) de l'écosystème marin pélagique. Une bonne représentation spatiale de cet environnement constitue donc un défi important.

Habituellement, la représentation et la visualisation de l'écosystème pélagique se fait par coupes statiques horizontales (Figure 1a) ou verticales (Figure 1b) (Carette, Mostafavi, et Devillers 2008; Head et al. 1997; Kaufman 1996). Dans les deux cas, la variabilité des phénomènes dans l'espace est représentée de manière continue au moyen d'une interpolation spatiale bidimensionnelle (2D) entre les points d'échantillonnage. Cela implique une perte d'informations spatiales puisque les jeux de données de l'écosystème pélagique décrivent l'espace 3D. L'étude de l'écosystème marin, ainsi que la gestion et la conservation des ressources marines, bénéficieraient grandement d'une représentation spatiale 3D adéquate.

La limitation au 2D de la représentation de l'écosystème pélagique peut s'expliquer par le manque d'outils adéquats pour la modélisation spatiale de cet environnement. Les Systèmes d'Information Géographiques (SIG) sont des outils d'observation, de représentation et d'analyse de premier plan pour la compréhension de données spatiales; donc un support nécessaire à la prise de décisions dans la gestion des zones côtières (Harris et al. 2008; Wright et Bartlett 2001). Or, les outils SIG ont principalement été développés pour des applications terrestres (Gold et Condal 1995; Ledoux et Gold 2008; Ng et al. 2007; Wright et Goodchild 1997) et ont jusqu'à ce jour été peu utilisés pour la modélisation spatiale 3D de l'environnement marin (Arsenault et al. 2004; Basu 1996; Carette, Mostafavi, et Devillers 2008; Gold et Condal 1995; Mesick et al. 2009; Ng et al. 2007; Wright et Goodchild 1997). En parallèle du développement des SIG, les outils de visualisation scientifique ont évolué, principalement sous l'impulsion de l'industrie pétrolière et gazière pour la modélisation du sous-sol. Bien que ces outils soient moins puissants que les SIG en ce qui concerne la gestion des données et leur facilité d'utilisation, ils sont utiles et permettent de traiter, au moins partiellement, le même type de champ continu 3D de variables que nous retrouvons dans l'écosystème pélagique.

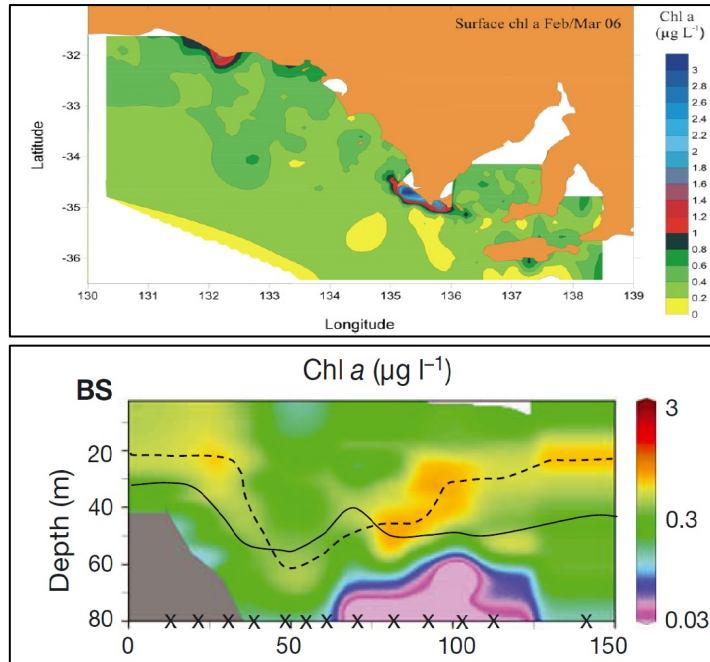


Figure 1. Représentation habituelle de phénomènes océanographiques. (a) Section horizontale représentant les variations spatiales de la concentration en chlorophylle a (chl a) (profondeur de 3 m) dans l'est de Great Australian Bight (Australie) (tiré de van Ruth, Ganf, et Ward 2010). (b) Section verticale de chl a dans la mer de Beaufort (tiré de (Martin et al. 2010)). L'axe horizontal de la Figure 1b correspond à la distance de la côte (km), alors que les marqueurs sur cet axe correspondent aux stations d'échantillonnage.

L'objectif général de ce travail de maitrise est d'explorer le potentiel des outils de modélisation 3D existants pour la représentation spatiale de l'écosystème pélagique marin. Plus particulièrement seront étudiés l'apport de la modélisation spatiale 3D et les performances des méthodes d'interpolation pour l'étude de l'écosystème pélagique du plateau Mackenzie dans le sud-est de la mer de Beaufort (océan Arctique).

1.2 Étude de cas - Mer de Beaufort, Arctique canadien

Un nombre croissant de scientifiques, gestionnaires et politiciens s'intéressent à l'environnement marin de l'Arctique, où les effets des changements climatiques s'accroissent et s'intensifient. Les modifications actuelles et appréhendées de l'écosystème Arctique pourraient influencer le processus de séquestration et de recyclage du carbone organique (Forest et al. 2013; Matsuoka et al. 2012; Sampei et al. 2011) concentré dans l'écosystème pélagique des plateaux continentaux des océans (Macdonald et al. 1998). Ce processus est fondamental pour la régularisation du climat de la planète et bien que la superficie de l'océan Arctique ne représente que 4 % de la superficie totale des océans, 20 % des plateaux continentaux à l'échelle mondiale s'y trouvent (Macdonald et al. 1998). Les principaux changements environnementaux observés dans l'océan Arctique découlent (Malina 2013):

- ✓ d'une diminution de couvert de glace, ce qui se traduit par une exposition accrue de l'écosystème pélagique à l'atmosphère ;
- ✓ d'une fonte progressive du permafrost combinée à l'augmentation du débit des fleuves, qui pourrait conduire à une augmentation de l'export de carbone organique terrigène vers l'océan Arctique;
- ✓ d'une augmentation de la radiation UV.

Dans ce contexte, la recherche scientifique dans l'océan Arctique s'est intensifiée depuis environ une décennie. Plusieurs campagnes océanographiques d'échantillonnage ont été organisées dans les eaux canadiennes (CASES en 2003-2004; ArcticNet-Malina en 2009) et à l'international (International Polar Year 2007-2008, <http://www.ipy.org/>). Les données utilisées dans le cadre du travail décrit ici proviennent de la campagne ArcticNet-Malina effectuée dans la partie sud-est de la mer de Beaufort, à bord du brise-glace de recherche canadien NGCC Amundsen. Cette région est caractérisée par un large plateau continental, le plateau Mackenzie, d'une superficie supérieure à 60 000 km² et délimitée par le Golfe d'Amundsen à l'est, le canyon du Mackenzie à l'ouest, et la mer de Beaufort au nord (Figure 2).

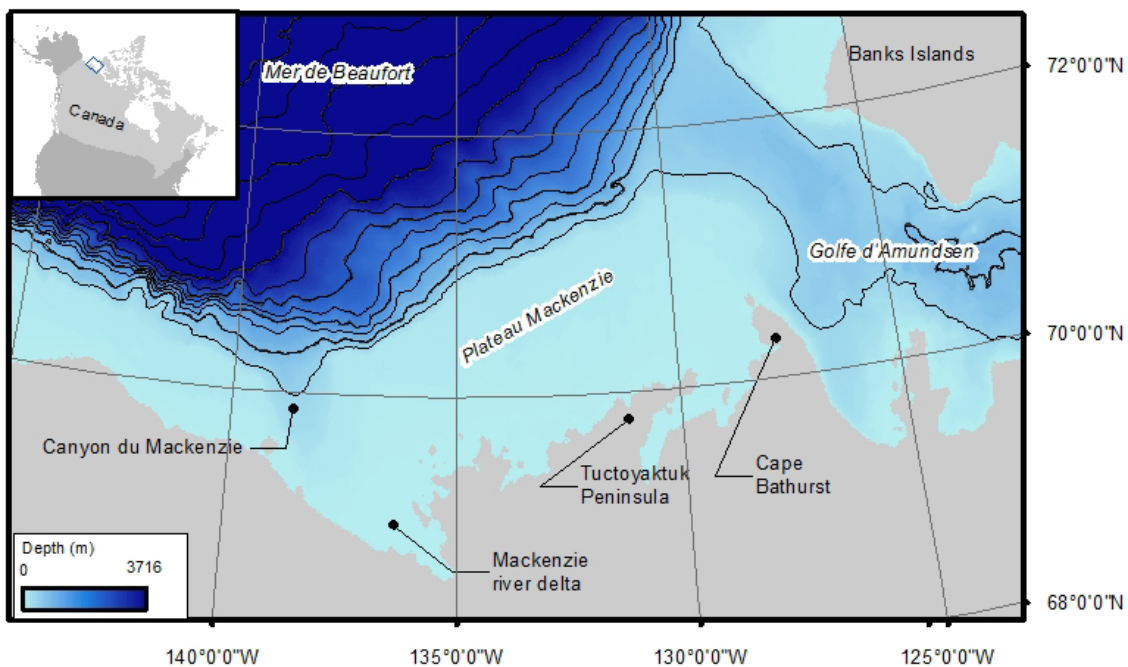


Figure 2. Région d'étude de la campagne océanographique ArcticNet-Malina (2009).

En ce qui concerne le travail présenté ici, la campagne océanographique ArcticNet-Malina se divise en deux parties distinctes. La première (leg 2A de l'expédition 2009 de l'Amundsen), effectuée du 16 au 26 juillet 2009,

a été chapeauté par le réseau de centres d'excellence du Canada ArcticNet. L'objectif de cet organisme, qui regroupe des scientifiques et des gestionnaires du milieu de la recherche, des gouvernements et des communautés nordiques, est *d'étudier les impacts des changements climatiques et de la modernisation dans l'Arctique canadien côtier* (ArcticNet 2010). La deuxième partie (leg 2B) de la campagne (*Malina*), fait partie d'un programme d'initiative française reposant sur une collaboration entre ArcticNet et des établissements français et américains de recherche. L'échantillonnage s'est effectué pendant 28 jours au mois d'août 2009 (suite au leg 2A).

Le programme Malina met l'emphase sur le réseau trophique inférieur (bactéries et plancton) et le régime physico-chimique qui en contraint le fonctionnement. Un des objectifs de Malina est de mieux comprendre comment la variabilité spatio-temporelle des facteurs abiotiques (ex. lumière, nutriments, couvert de glace, stratification, apports de substances fluviatiles) et des facteurs biotiques (ex. production primaire, abondance de phytoplancton, bactéries et zooplancton) qui influencent la dynamique des flux de carbone et d'azote organique particulaire (COP, CON) dans l'écosystème pélagique marin. Ces variables de l'écosystème pélagique sont donc directement liées aux taux de recyclage et de séquestration de carbone organique de l'océan Arctique.

De la campagne océanographique ArcticNet-Malina a résulté une vaste base de données spatiales comprenant plus d'une centaine de variables mesurées *in situ* ou dérivées de ces mesures. Ces données sont actuellement analysées par des équipes de chercheurs² multidisciplinaires et multinationales et constituent la base de l'expérimentation dans le cadre du travail de maîtrise décrit ici.

1.3 Problématique

1.3.1 Problématique générale

Habituellement, la représentation et la visualisation des variables continues de l'écosystème pélagique sont effectuées en 2D. Les formes fréquemment utilisées dans la littérature scientifique sont:

- ✓ la représentation 2D des variables dans le plan horizontal à une profondeur précise (souvent la surface de la mer) ou moyennée sur la colonne d'eau entière ou sur une partie de celle-ci;
- ✓ des coupes verticales de la colonne d'eau tirées d'un échantillonnage effectué le long d'une radiale.

² Lien des publications Malina : <http://malina.obs-vlfr.fr/publications.html>

Dans les deux cas, il s'agit de représentations qui ne considèrent que deux des trois dimensions de l'espace. Une analyse qui combine l'examen visuel de multiples coupes verticales et horizontales peut offrir une idée de la distribution 3D des phénomènes pélagiques d'un bassin océanique. Cependant, ce genre de représentation omet la continuité de l'analyse qu'offre le modèle 3D et peut impliquer un temps d'analyse considérable (Head et al. 1997; Kaufman 1996). Dans le cas d'une représentation par des coupes verticales, on ignore la variabilité entre les radiales d'échantillonnage souvent passablement éloignées les unes des autres. Par ailleurs, les technologies émergentes d'acquisition de données océanographiques impliquent une croissance des jeux de données multi-sources de distribution spatiale irrégulière entre les points de mesures. Ce type de données nécessite une modélisation spatiale 3D pour une représentation adéquate.

À la lumière de ces considérations, on constate que l'apport de la modélisation spatiale 3D serait important pour l'étude spatiale des phénomènes de l'écosystème pélagique marin. Les avantages de migrer vers une modélisation 3D de l'espace ont été évoqués dans d'autres domaines des géosciences : dans le cadre de la modélisation spatiale des sols (Cabello, Cuevas, et Ramos 2007; He et al. 2009) des nappes phréatiques (Raiber et al. 2012; Rivest, Marcotte, et Pasquier 2012; Robins et al. 2005), ou de la gestion des pêches (Carette, Mostafavi, et Devillers 2008), mais surtout lors par l'industrie des ressources minières qui utilise aujourd'hui ces méthodes de façon routinière (p.ex. Jones, McCaffrey et al. 2009).

1.3.2 Problématique spécifique

Les outils disponibles pour la modélisation spatiale 3D de l'écosystème pélagique marin sont actuellement peu nombreux, ce qui s'explique surtout par les contraintes imposées par la nature dynamique et tridimensionnelle de ce milieu (Gold et Condal 1995; Shyue et T sai 1996). En effet, plusieurs équipes de recherche ont reconnu l'absence d'outils de SIG marins (Arsenault et al. 2004; Mesick et al. 2009). Néanmoins, le développement récent de SIG et d'outils de visualisation scientifiques tendant vers la modélisation spatiale 3D a rendu disponible une gamme d'outils avec des fonctionnalités de modélisation qui ne sont pas encore (ou peu) exploités par les océanographes ou biologistes marins. Le potentiel de leur apport à l'étude de l'écosystème pélagique marin devrait faire l'objet d'un examen approfondi.

La qualité de la représentation spatiale dépend de plusieurs facteurs, dont la densité et la configuration de l'échantillonnage, ainsi que de la méthode d'interpolation (Gotway et al. 1996; Reed, Ellsworth, et Minsker 2004). L'interpolation spatiale, qui consiste à construire des champs de données continues à partir de données ponctuelles, est une étape essentielle de la modélisation spatiale des phénomènes continus de l'écosystème pélagique. De nombreuses études comparatives ont évalué la performance de méthodes d'interpolation spatiale en 2D (Li et Heap 2011; Li et al. 2011; Zimmerman et al. 1999). Toutefois, les études comparatives de méthodes d'interpolation 3D demeurent peu nombreuses et traitent principalement des

variables catégoriques³ (Falivene, Cabrera, et Saez 2007; He, Hu, Huang, et al. 2010). Aucune étude recensée ne traite de performance des méthodes d'interpolation 3D pour des phénomènes continus de l'environnement pélagique marin. Afin de pouvoir confirmer le potentiel des méthodes de modélisation spatiale 3D pour l'écosystème pélagique marin, il devient donc primordial de remédier à cette lacune.

1.4 Objectifs

L'objectif général de ce projet de recherche est de démontrer que la modélisation spatiale 3D (représentation et analyse) de l'écosystème pélagique marin constitue un apport important à la compréhension de son fonctionnement. Cet objectif se décline en trois sous-objectifs.

Dans un premier temps, je passe en revue les difficultés spécifiques que présente l'écosystème pélagique vis-à-vis de modélisation spatiale 3D et je décris les outils pertinents d'analyse de modèles 3D. J'effectue ensuite un inventaire de logiciels de modélisation spatiale 3D afin d'évaluer leur potentiel et limites pour la représentation, la visualisation et l'analyse 3D de l'écosystème pélagique marin.

Dans un deuxième temps, je vise à combler une lacune importante identifiée dans la littérature scientifique, à savoir la performance des méthodes d'interpolation 3D géostatistiques et déterministes pour des phénomènes continus de l'environnement pélagique marin. Ce sous-objectif représente, en termes d'efforts, la majeure partie de ce travail de maîtrise. Une méthode performante dans ce contexte réfère à

- une méthode d'interpolation *optimale*, soit une méthode qui crée une estimation la plus proche de la réalité que possible;
- une méthode d'interpolation *robuste*, soit une méthode qui est stable par rapport à une variation du paramétrage⁴.

Le critère 1 sera évalué par une estimation des erreurs pour chaque stratégie d'interpolation (méthode d'interpolation avec variation du paramétrage). Le critère 2 sera évalué par une variation du *nombre* et du *poids* de points de mesures distants utilisés dans l'interpolation. Deux des méthodes d'interpolation les plus couramment utilisées pour la modélisation spatiale 2D seront comparées: la méthode déterministe *distances inverses* (IDW) et la méthode géostatistique de *krigeage ordinaire* (OK).

³ Variable qui désigne et se rapporte à des attributs non quantifiables, aussi appelé variable qualitative.

⁴ Opération qui consiste à déterminer la valeur des paramètres contrôlant le comportement d'un logiciel, de manière qu'il puisse effectuer certains traitements, selon le contexte dans lequel on entend l'utiliser (Définition OQLF, 2013).

Je tenterai d'identifier une stratégie d'interpolation optimale et robuste pour cinq phénomènes pélagiques (variables) d'une complexité croissante en termes d'anisotropie de données, de configuration spatiale d'échantillonnage et de fonction écologique (le lien avec les cycles des éléments nutritifs). Ceci permettra donc de valider la performance de la méthode d'interpolation 3D à plusieurs niveaux de complexité (c.-à-d., la performance de la même méthode d'interpolation est évaluée pour plusieurs variables). Les phénomènes pélagiques pour lesquelles la modélisation est proposée – température (T), concentration en chlorophylle a (chl a), coefficient d'atténuation des particules en suspension (c_p), distribution des eaux de l'halocline supérieure (UHW) et flux vertical de carbone organique particulaire (flux de COP) – sont issus d'une campagne d'échantillonnage typique de la recherche océanographique.

Finalement, un dernier sous-objectif dans le cadre de ces travaux est l'application des méthodes d'interpolation 3D aux variables pélagiques issues de la campagne océanographique ArcticNet-Malina. Le potentiel et la pertinence de la représentation 3D seront ainsi vérifiés par une validation qualitative de l'aspect visuel des modèles spatiaux 3D.

Ces travaux se limitent à la représentation et distribution spatiale des phénomènes continus dans l'espace 3D, tel que des variables physico-chimiques et biologiques (e.g. T, chl a etc.). L'étude des phénomènes continus représente le principal objet de la recherche océanographique, mais également des géosciences connexes (météorologie, géologie, géophysique, etc.). La représentation des *objets* (discrets) dans cet environnement (poisson, mammifères marins, etc.) constitue un défi géomatique bien distinct qui ne sera pas abordé.

Il importe de souligner que dans le cadre de cette recherche, je traite de modélisation statistique destinée à produire par interpolation 3D des champs de données continus à partir d'échantillons *in situ* discrets. Je ne fais pas référence à la modélisation numérique largement répandue en océanographie pour prédire la dynamique de l'océan et les processus biogéochimiques.

Il est également important de souligner que l'aspect temporel (dynamique) de l'environnement pélagique est peu abordé dans ce mémoire. Lorsqu'on souhaite modéliser l'aspect dynamique dans une représentation 3D, on ajoute une contrainte importante en ce qui concerne la structure de données (Arsenault et al. 2004; Ledoux et Gold 2006; Ledoux et Gold 2008). Toutefois, le focus de ce mémoire est sur la l'aspect spatial 3D et l'influence du facteur temporel est plutôt abordé en relativisant les résultats à l'influence du temps écoulé entre les mesures utilisées pour la modélisation.

1.5 Méthodologie

La méthode de recherche adoptée pour répondre aux objectifs énoncés dans la précédente section est composée de plusieurs étapes illustrées dans la Figure 3 qui se répartissent entre recherche théorique (recensement de la littérature, analyse des besoins, inventaire des outils) et recherche expérimentale (inventaire et préparation des jeux de données, plan expérimental, interpolation et modélisation, analyses statistiques).

L'objectif du recensement la littérature est de placer le projet dans le contexte actuel de la recherche océanographique, de cibler des enjeux réels et d'identifier les besoins en termes de représentation spatiale de l'écosystème pélagique marin. Ces besoins ont été validés entre autres dans le cadre d'une collaboration entre l'équipe de recherche du prof. Mir Abolfazl Mostafavi, Directeur du Centre de recherche en géomatique, et des chercheurs (dont la contribution significative d'Alexandre Forest) de l'équipe de Marcel Babin, titulaire de la Chaire d'excellence en recherche du Canada sur la télédétection de la nouvelle frontière arctique du Canada au département de la biologie de l'Université Laval.

Le deuxième volet du recensement de la littérature visait à explorer le potentiel des outils existants de modélisation spatiale 3D pour l'écosystème pélagique marin. Les avantages, étendues et limites des différents outils SIG de visualisation scientifique, commerciaux et open source ont fait l'objet de cet inventaire. L'un des attendus de cette phase était de cibler quelques outils prometteurs pour utilisation ultérieure, dans la phase expérimentale des jeux de données. Ce deuxième volet constitue donc la démarche visant à répondre au premier sous-objectif stipulé dans la section 1.3.

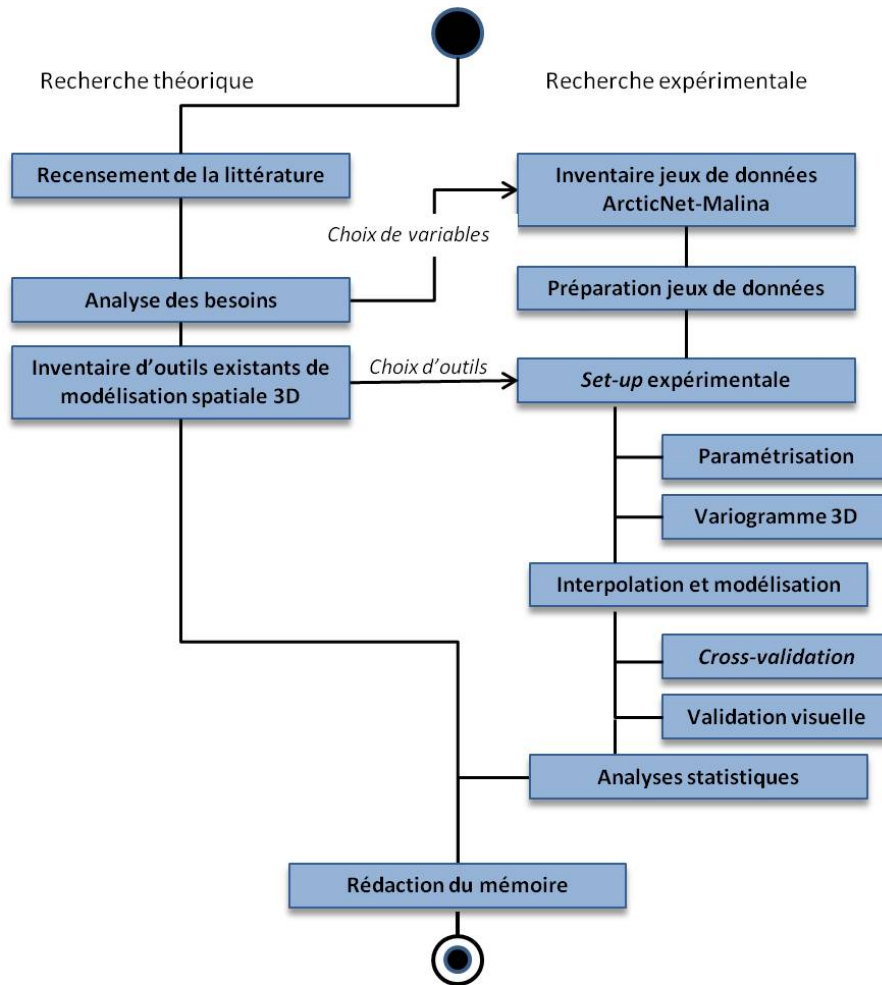


Figure 3. Méthode pour la réalisation du projet de recherche.

La phase théorique de mes travaux visait également à cibler quelques variables, issues de la campagne océanographique ArcticNet-Malina, à utiliser dans la phase expérimentale. Ces variables (T, chl *a*, c_p , UHW, flux de POC) ont été choisies en fonction de leur importance dans la recherche océanographique en général, de leur pertinence en lien avec les objectifs de ArcticNet-Malina (rôle de l'océan Arctique dans la régularisation du climat à l'échelle globale) et en fonction de leur complexité spatio-temporelle, mais également en raison de la densité de données disponibles. Alors que la T, le c_p , et le chl *a* constituent des variables mesurées par défaut dans toute campagne de recherche océanographique, l'UHW et le flux de POC représentent des variables dérivées de méthodes de calculs complexes.

La phase expérimentale s'est amorcée par les étapes *d'inventaire* et de *préparation des jeux de données*, ce qui comprenait (entre autres) le « nettoyage » des données, l'exploration de la distribution spatiale des valeurs et de la configuration de l'échantillonnage des jeux de données en ArcGIS 10, l'identification des valeurs

extrêmes, de la modélisation de variogramme 3D, etc. Des efforts considérables ont également été consentis à se familiariser avec outils de modélisation choisis (Paradigm GOCAD® ver. 2009.2, et EnterVol for ArcGIS ver. 1.4, de CTech) afin d'optimiser les interpolations et les modélisations. De multiples échanges ont également eu lieu avec les développeurs de ces outils afin de permettre leur exploitation maximale dans le cas spécifique de la modélisation spatiale de l'écosystème pélagique marin.

L'approche utilisée lors de l'évaluation des performances des méthodes d'interpolation et de modélisation 3D (sous-objectifs 2 à 3) est présentée dans la section *Methods* de l'article 2 (Chapitre 4) et ne sera donc pas répétée dans la présente section.

Je tiens finalement à préciser mon implication dans un article publié en 2013 avec le titre *Ecosystem function and particle flux dynamics across the Mackenzie Shelf (Beaufort Sea, Arctic Ocean): an integrative analysis of spatial variability and biophysical forcings* (Forest et al. 2013). Cette implication consistait en la création d'une représentation 3D du flux de COP échantillonnée durant la campagne d'ArcticNet-Malina et était donc un fruit des acquis des travaux de recherche,

1.6 Structure du mémoire

Ce mémoire est divisé en cinq chapitres, dont deux sous forme d'articles scientifiques. Le premier chapitre introduit le contexte de la recherche, l'étude de cas de la campagne océanographique ArcticNet-Malina, la problématique adressée, ainsi que les hypothèses posées et les objectifs poursuivis.

Le Chapitre 2 constitue une revue des concepts liés à la modélisation et interpolation spatiale 3D de manière générale, ainsi qu'au contexte spécifique de l'écosystème pélagique marin. Les caractéristiques particulières de l'écosystème pélagique d'un point de vue de la géomatique et les contraintes imposées par cet environnement sont également abordées dans ce deuxième chapitre.

Le Chapitre 3 présente les résultats de l'inventaire des outils de modélisation spatiale 3D et discute leurs limites et potentiels dans le contexte de la modélisation spatiale de l'environnement marin. Ce chapitre est structuré sous forme d'article et constitue en fait un résumé long, accepté à la conférence du 7th 3DGeoInfo (avec comité de lecture), tenue à Québec au mois de mai 2012. Sa référence est la suivante :

Sahlin J., Mostafavi M. A., Forest A., Babin M. et Lansard B. (2012). 3D geospatial modelling and visualization for marine environment: Study of the marine pelagic ecosystem of the south-eastern Beaufort Sea, Canadian Arctic. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Québec, ISPRS Archives, Copernicus Publications. Volume XXXVIII-4/C26, 21-24. DOI : 10.5194/isprsarchives-XXXVIII-4-C26-21-2012

Le Chapitre 4 constitue le cœur de ce mémoire et évalue le potentiel des méthodes d'interpolations et de modélisation spatiale 3D pour la représentation des phénomènes continus de l'écosystème pélagique marin de la mer de Beaufort (Arctique Canadien). Ce chapitre est également sous forme d'article, accepté pour publication dans la revue scientifique *Marine Geodesy* et sa référence temporaire est la suivante :

Sahlin J., Mostafavi M. A., Forest A., Babin M. (2014). *Accepté pour publication*. Assessment of 3D spatial interpolation methods for study of the marine pelagic environment. *Marine Geodesy*, special issue on Marine and Coastal Geographical Information Systems (GIS) in Vol. 37, No. 2, 2014.

Les deux articles ont été rédigés en anglais avec des résumés en français. Le Chapitre 5 conclut ce mémoire avec un retour sur les objectifs présentés dans le Chapitre 1.

CHAPITRE 2 - REVUE DES CONCEPTS

Dans ce chapitre, je définis premièrement l'écosystème pélagique marin et en décris les caractéristiques particulières contraignantes pour la modélisation spatiale (section 2.1). La section suivante (2.2), constitue une revue des concepts liés à la modélisation et l'interpolation spatiale 3D de manière générale, ainsi qu'au contexte spécifique de l'écosystème pélagique marin.

2.1 L'écosystème pélagique marin

2.1.1 Délimitation

Ce mémoire reprend la définition de Game, Grantham et al. (2009) de l'écosystème pélagique marin : des phénomènes physiques, chimiques ou biologiques de la colonne d'eau des océans ou de la pleine mer, plutôt que des eaux adjacentes au milieu terrestre ou au milieu côtier (traduction libre). Ainsi, l'écosystème pélagique marin consiste en l'espace volumétrique délimité par la surface et le fond de l'océan ou du bassin marin.

La partie expérimentale de ce travail de maîtrise se concentre plus particulièrement sur l'écosystème pélagique marin de la mer de Beaufort, mais les résultats et la discussion associée peuvent s'étendre au contexte de la modélisation spatiale 3D de tout espace aquatique (bassins océaniques, milieux côtiers, grands lacs). Certaines caractéristiques de l'écosystème pélagique marin (d'un point de vue géomatique) sont partagées avec celles d'autres milieux (p. ex. nature volumétrique, dynamique), alors que d'autres caractéristiques sont plutôt spécifiques à l'écosystème pélagique marin tels que l'échelle spatiale, la configuration spatiale des jeux de données, etc. (voir la section suivante). Ces dernières imposent des contraintes au développement de modèles spatiaux, qui seront abordées dans les trois sections suivantes. Il est à noter qu'une revue approfondie des forçages océaniques et processus physico-chimiques qui dictent la variabilité spatio-temporelle des phénomènes pélagiques marins va au-delà du périmètre de ce travail de maîtrise.

2.1.2 Forçages océaniques et variabilité spatio-temporelle

Les variables fondamentales de l'écosystème pélagique des océans (p. ex. salinité, température) sont relativement stables dans le temps et leur échelle spatiale sur le plan horizontal : planétaire, méso-échelle (~2 à 300 km) ou sub-méso-échelle (< 2 km). Sur la verticale, l'échelle des variations spatiales est beaucoup plus petite. Un changement de valeur observé au plan horizontal sur plusieurs dizaines de kilomètres peut être observé sur quelques mètres en fonction de la profondeur (Figure 4). Cette stratification horizontale des variables constitue un trait principal de l'écosystème pélagique marin.

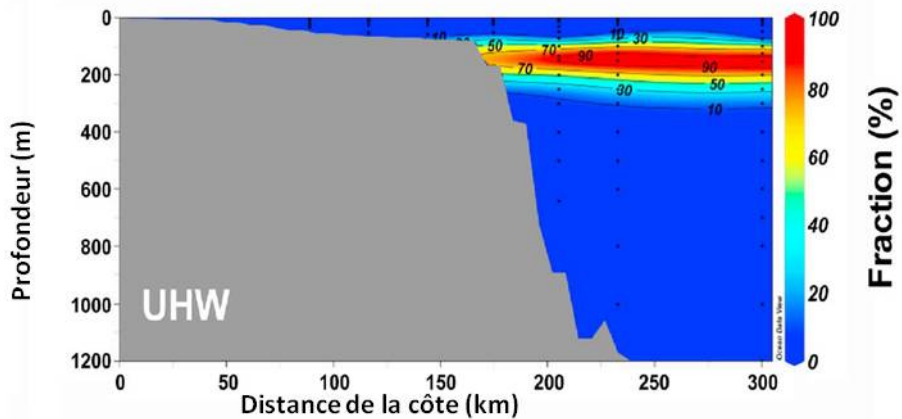


Figure 4. Exemple de stratification dans l'écosystème pélagique marin. Coupe verticale de la masse d'eau de l'halocline supérieure (UHW) dans la mer de Beaufort. Le dégradé de couleurs représente la fraction de l'écosystème pélagique constituée de l'UHW. La zone grise représente le plateau continental. La figure est tirée de Lansard, Mucci et al. (2012).

La répartition spatiale de la salinité et de la température, dont les variations sont fortement liées, est un autre aspect fondamental des océans. Leurs maxima et minima sont utilisés pour caractériser et identifier des masses d'eau d'origine distincte (p.ex. eaux de surface de l'Atlantique, eaux de l'halocline supérieure d'origine Pacifique, etc.), avec des frontières plus au moins distinctes (e.g. Lansard, Mucci et al. 2012). Leur répartition (et par conséquent la distribution spatiale d'autres variables) est influencée par différents forçages océaniques : les grands patrons de circulation de courants profonds et de courants de surface, les échanges de chaleurs et d'eau douce avec l'atmosphère et les fleuves, ainsi que les variations de la pression atmosphérique et de gravité.

À plus petite échelle (mésos- ou submésos-échelle), des événements tels des remontées d'eaux profondes, des facteurs tels que la distance à la côte, la proximité de fleuves, ou la présence de glace de mer influencent la variabilité spatiale et temporelle des caractéristiques de l'océan et des phénomènes qu'on y observe. La compréhension de cette variabilité est fondamentale dans l'étude de la productivité des océans et par conséquent, du cycle de carbone de la planète. Les événements de remontée d'eaux profondes amènent des eaux riches en nutriments des fonds vers la couche euphotique (zone de pénétration de la lumière⁵) et stimulent ainsi la production primaire⁶. La présence de rivières peut constituer une source additionnelle d'éléments nutritifs. Dans la région d'étude des présents travaux, la rivière Mackenzie est responsable d'une décharge annuelle d'environ $\sim 330 \text{ km}^3$ d'eau douce riche en certains nutriments (Malina 2013). Dans les

⁵ Couche de la surface de la mer délimitée par la profondeur à laquelle on mesure 1 % de la lumière du Soleil pénétrant sous la surface ((Minster 1997).

⁶ Quantité totale de matière organique fixée dans la photosynthèse chlorophyllienne (bactéries et plantes).

régions polaires, la couverture de glace pendant plusieurs mois de l'année est un autre facteur influençant la productivité des océans.

La variabilité spatio-temporelle des phénomènes de l'écosystème pélagique marin repose donc sur des processus océanographiques et côtiers complexes, qui agissent sur une échelle variant du local au planétaire. L'étude de la répartition spatiale de ces phénomènes fait partie de la grande quête de compréhension du fonctionnement des océans et du climat planétaire. Inversement, ces processus océanographiques imposent des contraintes particulières à la modélisation spatiale de l'écosystème pélagique.

2.1.3 Caractéristiques particulières d'un point de vue de la géomatique

L'environnement pélagique marin est caractérisés par sa nature continue, dynamique et volumétrique, ainsi que par ses limites floues (Carette, Mostafavi, et Devillers 2008; Gold et Condal 1995; Ledoux et Gold 2006; Shyue et Tsai 1996).

L'écosystème pélagique marin est un vaste continuum de phénomènes biotiques et abiotiques dans lequel on note la présence sporadique d'objets distincts, tels que des poissons ou des mammifères marins. D'un point de vue conceptuel, l'écosystème pélagique marin est donc sensiblement différent de l'écosystème terrestre, souvent considéré comme un espace vide rempli d'objets (maisons, arbres, etc). Dans la recherche océanographique de l'écosystème pélagique, l'étude spatiale et temporelle des phénomènes se réfère surtout à des phénomènes continus, ce qui est également le cas pour des domaines des géosciences connexes tels que la météorologie, la géologie et la géophysique (Ledoux and Gold 2006). La distinction entre des phénomènes continus (*champs*) et des phénomènes *discrets* reflète deux visions conceptuelles de l'espace qui impose des défis distincts de modélisation spatiale (voir section 2.2).

La caractéristique 3D de l'écosystème pélagique marin ajoute une contrainte importante à sa représentation, son analyse et sa visualisation (Coley et Claburn 2005; Kaufman 1996; Kracker 1999; Ledoux et Gold 2008). Cette contrainte s'impose au niveau de la conception et de l'implantation de la structure des données 3D (DeMers 2002), mais également au niveau de la gestion et du stockage des données. Les SIG, qui sont des outils d'observation, de représentation et d'analyse de premier plan pour la compréhension de données spatiales, ont jusqu'à récemment été très limités pour la modélisation spatiale 3D (Arsenault et al. 2004; Basu 1996; Carette, Mostafavi, et Devillers 2008; Gold et Condal 1995; Mesick et al. 2009; Ng et al. 2007; Wright et Goodchild 1997). En outre, le développement récent des fonctionnalités 3D a été dirigé vers des applications terrestres et pour la représentation de phénomènes discrets. Tel que mentionné précédemment, la représentation des phénomènes pélagiques devrait comprendre les phénomènes 3D continus. Selon DeMers

(2002) les structures de données représentant des phénomènes continus étaient encore au stade expérimental au début des années 2000.

Le caractère *dynamique* de la colonne d'eau impose qu'on considère le facteur temporel lors de l'interprétation et l'analyse des jeux de données océanographiques. Contrairement à d'autres univers volumétriques des sciences de la terre (notamment le sous-sol géologique qui peut être considéré relativement constant), la colonne d'eau est en perpétuel changement. Lorsqu'on souhaite modéliser l'aspect dynamique dans une représentation 3D, on ajoute encore une fois une contrainte importante en ce qui concerne la structure de données (Arsenault et al. 2004; Ledoux et Gold 2006; Ledoux et Gold 2008). La constance des données océanographiques, souvent présumée lors d'une modélisation spatiale, est donc à strictement parler toujours une approximation plus au moins bonne de la réalité. En raison des grands espaces et des contraintes au plan de l'acquisition de données (voir section suivante), plusieurs semaines peuvent séparer des points d'échantillonnage lors d'une campagne océanographique. Toutefois, comme nous l'avons mentionné précédemment, les paramètres fondamentaux de l'océan sont relativement stables dans le temps (Minster 1997) et lorsque la variabilité des phénomènes est discutée dans le contexte du potentiel d'influence du temps écoulé entre les mesures, cette absence de constance *stricto sensu* peut être relativisée. Le date de l'échantillonnage n'est qu'une parmi d'autres facteurs qui influence des variables pélagiques; dans le cas spécifique de la campagne ArcticNet-Malina, ce facteur n'expliquait que 4,7% de la variation totale du flux verticale de POC (Forest, Babin et al. 2013).

La caractéristique des *limites floues* des phénomènes fait référence à des situations où la différenciation entre des éléments géographiques n'est pas nette, mais plutôt progressive (Carette, Mostafavi, et Devillers 2008). Dans l'écosystème pélagique marin, un banc de poissons et un nuage de plancton constituent deux exemples des phénomènes avec des limites floues. La complexité de la représentation des phénomènes avec des limites floues constitue l'objet de plusieurs travaux récents (e.g.(Altman 1994; Carette, Mostafavi, et Devillers 2008; Zheng et al. 2012))

À ces contraintes, on devra également ajouter la particularité de l'échelle spatiale dans la recherche océanographique. À titre d'exemple, des agrégations de zooplancton peuvent être identifiées sur une échelle de quelques millimètres à presque 100 km (Bathmann et al. 2001). Le choix de la résolution spatiale utilisée dans la représentation des données est dans ce cas critique afin d'atteindre le degré de détails visé par l'étude. Mais avant tout, la précision de la représentation spatiale dépend de la densité et de la disponibilité des données océanographiques, ce qui est restreint, par la difficulté que présente l'acquisition des données dans le milieu marin.

2.1.4 Méthodes d'acquisition de données

En milieu marin, l'acquisition de données est difficile et implique des ressources financières et humaines considérables. Depuis les débuts de la recherche océanographique, la collecte de données sur l'écosystème pélagique est effectuée lors de campagnes océanographiques à l'aide d'équipements largués des navires. Parmi les instruments largement utilisés dans ce contexte, on retrouve des CTD (Conductivity, Température, Depth) et des rosettes (qui permettent le prélèvement d'eau profonde dans des bouteilles au moyen d'un déclenchement à partir de la surface). Les avancements technologiques des dernières décennies ont rendu accessible une vaste gamme de technologies alternatives permettant l'acquisition de données. La plus marquante est peut-être la télédétection par satellite, qui permet de fournir des données globales, continues et répétées de la surface de l'océan (Minster 1997). Parmi les variables étudiées à partir de l'espace, on peut mentionner la topographie de la surface de la mer (*altimétrie satellitaire*), la température (mesure du flux infrarouge), la couleur de la mer (qui permet une estimation de la production planctonique) et la présence de vagues et de glaces de mer (diffusiométrie). En revanche, les données satellites ne s'appliquent qu'à la couche de surface des océans (quelques millimètres à quelques dizaines de mètres) et doivent par conséquent être complétées par des mesures *in situ*.

La télédétection acoustique à l'aide de sonars déployés *in situ* permet d'obtenir la distribution spatiale des ressources biologiques et physiques dans la colonne d'eau. Un sonar (acronyme de *SOund NAVigation and Ranging*) est un appareil qui utilise les propriétés de la propagation du son dans l'eau pour détecter et situer des objets sous l'eau. L'onde acoustique est dispersée, absorbée et reflétée par les objets et la mesure de la force et du temps de retour de signal peuvent donc être utilisées pour déterminer la localisation et la masse de l'objet et même sa taille (p. ex., poissons). Puisque la dispersion du son est fonction de la température, de la salinité et de la profondeur, les sonars acoustiques peuvent être utilisés pour déduire des propriétés physiques de la colonne d'eau.

À ces technologies, on peut ajouter le mouillage (instrument à poste fixe), de bouées dérivantes (flotteurs en surface ou en profondeur qui mesurent les courants lagrangiennes⁷), et de planeurs sous-marins téléguidés. Ces technologies d'acquisition présentent toutes des limites et des défis géomatiques sur différents plans. On note premièrement la très grande quantité d'informations générées par la télédétection (satellite et sonars). Des efforts considérables sont nécessaires pour rendre cette information disponible et compréhensible, non seulement pour les gestionnaires, mais également pour les scientifiques. En ce qui concerne les planeurs sous-marins ou les flotteurs, on gère des mesures ponctuelles et spatialement dispersées. En revanche, les

⁷ Mesure en dérive : suivi du mouvement des particules.

techniques d'acquisition *in situ* produisent des jeux de données avec une configuration spatiale des points abondante sur le plan vertical, mais peu dense sur le plan horizontal. La distribution non homogène de l'échantillonnage en mer est également accentuée par la configuration de l'échantillonnage des campagnes océanographiques, souvent effectués le long de radiales (Forest et al. 2013; Kaufman 1996; Lansard et al. 2012; Martin et al. 2010; van Ruth, Ganf, et Ward 2010).

Les particularités de l'acquisition de données de l'écosystème pélagique marin impliquent donc des contraintes à l'égard de la disponibilité et de la répartition de données spatiales. Ces contraintes affecteront, à des niveaux différents, la qualité de la modélisation spatiale (Gotway et al. 1996; Reed, Ellsworth, et Minsker 2004). Pour bien saisir les limites, ainsi que le potentiel de la modélisation spatiale 3D pour l'écosystème pélagique, il faut comprendre les concepts de base de la modélisation spatiale, ce qui constitue l'objet des prochaines sections.

Il est finalement à noter que la nature (distribution spatiale) des jeux de données de l'écosystème pélagique ressemble beaucoup à ce qu'on traite lors de l'étude du sous-sol géologique- données très denses à la verticale et très espacées à l'horizontale. Dans le milieu marin, les stations horizontales d'échantillonnage CTD peuvent ainsi être comparées à des trous de forage (*wells*) des disciplines de géologie et de génie minier. Cette comparaison est pertinente alors que plusieurs outils sophistiqués de visualisation scientifique ont été développés pour la représentation 3D dans le contexte du génie minier.

2.2 Modélisation spatiale 3D de l'écosystème pélagique

2.2.1 Conceptualisation de l'espace et modélisation spatiale

Dans ce mémoire, la *modélisation spatiale* réfère à l'ensemble des étapes nécessaires à la représentation (dans un environnement numérique) d'un phénomène géographique (phénomène géométrique). Le *modèle spatial* est ainsi utilisé pour représenter le phénomène géographique et pour déduire de l'information sur la relation spatiale entre des phénomènes.

La création du modèle spatial peut être divisée en trois niveaux d'abstraction distincts (Figure 5), soit *le modèle conceptuel*, *le modèle de données* et *la structure de données* (Burrough et McDonnell 1998; Hashemi Beni 2009; Peuquet 1988). Ces niveaux suivent un ordre d'abstraction décroissant et du moins au plus détaillés (Peuquet 1988). Ils permettent de passer d'une vision conceptuelle de la réalité vers l'implantation d'un modèle de données qui permet la représentation numérique des phénomènes géographiques.

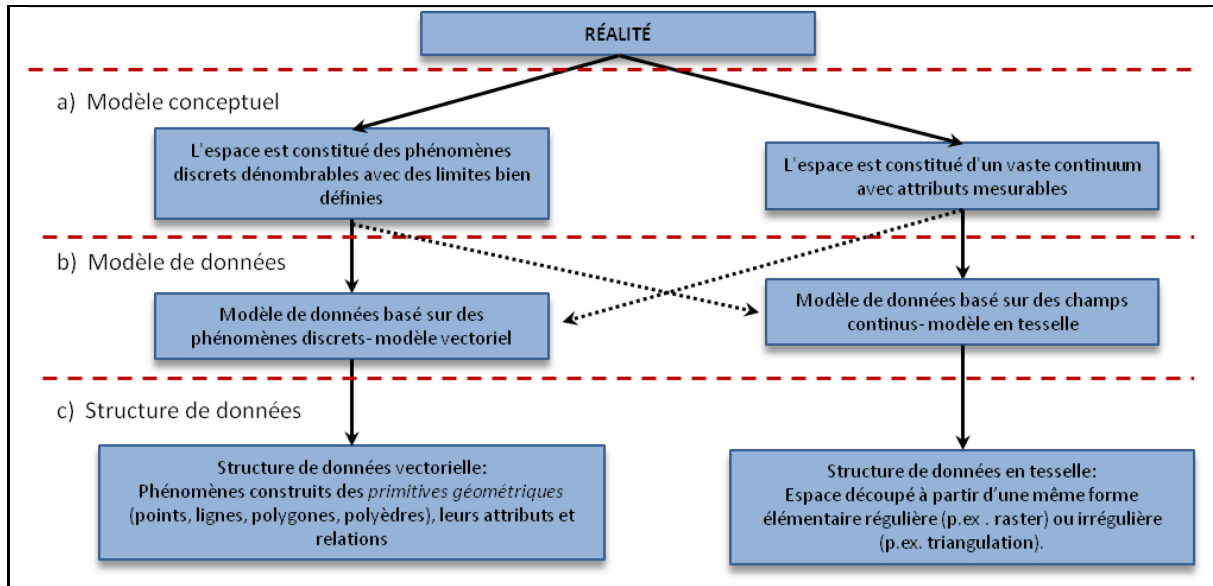


Figure 5. Étapes pour la création du modèle spatial de phénomènes géographiques (adapté de Burrough et Macdonnell 1998; Hashemi Beni 2009).

Sur le niveau conceptuel, la distinction entre des phénomènes (objets⁸) *discrets* et des phénomènes continus (*champs*) reflète deux visions de l'espace (Burrough et Mcdonnell 1998; Goodchild 1992; Hashemi Beni 2009). Dans le premier cas, l'espace est considéré vide, rempli de phénomènes discrets et dénombrables avec des limites bien définies. Dans le deuxième cas, l'espace est constitué d'un vaste continuum avec des attributs mesurables pour chaque localisation échantillonnée.

Le deuxième niveau d'abstraction correspond à la création du *modèle de données* qui peut être défini comme [...] *the formalized equivalents of the conceptual models used by people to perceive geographical phenomena* (Burrough et Mcdonnell 1998). Ainsi, le modèle de données formalise la façon dont l'espace est découpé, il suppose que les phénomènes peuvent être mesurés ou spécifiés et que les coordonnées géographiques peuvent être précisées (*idem*).

Finalement, *la structure de données* spatiale constitue le dernier niveau d'abstraction de la modélisation spatiale. Elle réfère à l'implantation du modèle de données et fournit par conséquent l'information dont l'ordinateur a besoin pour reconstruire le modèle de données dans son format informatique (Hashemi Beni 2009). Un modèle de données peut ainsi être implanté avec plusieurs structures de données différentes (Ledoux 2006).

⁸ Si un objet spatial peut référer aussi bien à un phénomène continu qu'à un phénomène discret, on préfère toutefois, dans ce mémoire, utiliser le terme phénomène, puisque le terme objet fait intuitivement référence à une structure vectorielle (vue conceptuelle orientée-objet).

2.2.1.1 Modèles de données et structures de données

Les deux visions conceptuelles de l'espace ont abouti à deux approches fondamentales de modèles de données géométriques, soit le modèle **vectorel** et le modèle **en tesselle** (Burrough et McDonnell 1998; Goodchild 1992; Hashemi Beni 2009; Mostafavi 2002). Dans le modèle vectorel, on représente des objets spatiaux en les transformant en des *primitives géométriques*. Celles-ci correspondent habituellement aux points, lignes, polygones et polyèdres (p. ex., un point et un polygone peuvent respectivement représenter une ville et un lac). Les primitives géométriques sont présentées dans la structure de données vectorielle correspondante par des séries de coordonnées et par leur relation spatiale (Burrough et McDonnell 1998; Cromley 1992). Les structures de données vectorielles 3D comprennent, entre autres, le modèle par frontières (B-Rep) et les modèles de *constructive solid geometry* (CSG) (Pouliot, Lachance, et Kirkwood 2006; Shyue et Tsai 1996).

Le modèle de données la plus fréquemment utilisée pour la représentation de variables continues est le modèle en tesselle (Burrough et McDonnell 1998). Celle-ci découpe l'espace à partir d'une même forme élémentaire régulière (p.ex. cellules carrées, triangles, cubes) ou irrégulière (p.ex. triangulation de Delaunay et le diagramme Voronoi). La structure *raster* est une des structures en tesselle (régulière) la plus répandue lors de la représentation 2D. L'espace est découpé en une grille de cellules carrées (*pixels*) dont l'emplacement est défini par le numéro de sa colonne et de sa rangée (DeMers 2002). Le même principe peut être utilisé pour la représentation de l'espace 3D et on désigne alors les cellules *voxel* (VOLumetric piXEL; Burrough et McDonnell 1998). Parmi les structures de données 3D en tesselle, on note la structure de *3D Array* et la structure *Octree* (Li 1994).

Il existe un certain chevauchement entre les modèles de données basés sur les phénomènes discrets ou continus et leurs structures de données associées (illustré dans la Figure 5 par les flèches pointillées). Ainsi, un phénomène continu, telle que la profondeur de l'eau, est habituellement représenté par une structure en tesselle, alors que l'utilisation d'une structure vectorielle par des *isolignes* (voir section 2.2.5), (qui relie des séries de coordonnées aux attributs de valeur égale), constitue une structure alternative ou complémentaire. La Figure 2 combine notamment ces deux structures de données pour représenter la profondeur d'eau dans la mer de Beaufort.

Le modèle de données en tesselle régulière présente plusieurs avantages par rapport aux modèles vectoriels. Notamment, il est simple à comprendre et à programmer et certaines relations spatiales (p.ex. superposition) sont faciles à déduire. Toutefois, le modèle en tesselle nécessite beaucoup de mémoire et ne permet pas de préserver l'intégrité des objets : il s'agit d'une approximation des objets, traduits en pixels ou voxels (Burrough

et McDonnell 1998; Cromley 1992; Pouliot et al. 2008). La précision de la représentation du phénomène spatial sera ainsi en fonction de la taille des cellules utilisées pour le modèle.

2.2.2 Modélisation spatiale 3D des variables continues

L'espace de modélisation le plus souvent utilisé pour la cartographie de l'univers est constitué d'une carte ou d'un plan 2D. La localisation d'un point dans cet univers est définie par deux coordonnées (x, y). Une *modélisation spatiale 3D* sous-entend que chaque point dans l'espace est défini par trois coordonnées (x, y, z) avec au moins un attribut qui définit une valeur pour chaque point (van der Putt 2009). Pour représenter des phénomènes 3D, on utilise un univers 3D qui exploite les axes X, Y et Z.

Dans le langage commun de plusieurs SIG, on utilise le terme de *modélisation 3D* dans le contexte de construction de modèles numériques de terrain (MNT). Dans ce dernier cas, le positionnement vertical d'un phénomène spatial (discret ou continu) est indiqué par une valeur d'attribut par rapport à une surface de référence (p.ex. le niveau moyen des mers), plutôt que par une troisième coordonnée. Ce genre de représentation devrait plutôt être considéré comme une modélisation 2,5D (Tse et Gold 2004), et être défini comme *une représentation simplifiée de la troisième dimension* (par définition : OQLF 2013). Les TIN (*Triangulated Irregular Network*), couramment utilisés pour représenter la topographie des surfaces, constituent des exemples de modélisations spatiales 2,5D. Dans une *vraie* modélisation spatiale 3D, la même localisation (x,y) peut avoir plusieurs valeurs de z, ce qui nous permet ainsi de décrire l'espace entre ces valeurs de z.

Les coupes verticales de la colonne d'eau fréquemment utilisées pour représenter la distribution spatiale des variables de l'écosystème pélagique (section 1.1) constituent un cas particulier d'une représentation 2D. Ces coupes peuvent être considérées comme des cartes 2D avec un système de coordonnées locales, où l'axe X représente la distance parcourue depuis le début du transect d'échantillonnage et l'axe Y représente la profondeur de l'échantillonnage. On présume habituellement une linéarité du transect, car les stations d'échantillonnage à l'horizontal sont projetées sur cette ligne avant l'interpolation spatiale⁹. Cette modélisation spatiale est donc très différente d'une coupe verticale dans un modèle volumétrique, alors que dans ce dernier cas le composant 3D des données spatiales a été préservé.

Le problème spécifique de la représentation des phénomènes continus découle du fait qu'il est impossible d'échantillonner des variables partout dans l'espace (Ledoux et Gold 2008; Li et Heap 2008). On doit par

⁹ Cette méthode est utilisée par l'outil Ocean Data View (ODV), généralement utilisé par les océanographes pour la représentation spatiale des phénomènes continus de l'écosystème pélagique marin.

conséquent reconstruire ces champs à partir de données ponctuelles. Cette reconstruction s'effectue par une *interpolation spatiale*, qui constitue donc une étape essentielle de la représentation spatiale de phénomènes continus par une structure raster 2D ou 3D. L'interpolation spatiale sera décrite dans la section 2.2.4 et fera également l'objet du chapitre 4.

2.2.2.1 Raster 3D

Tel que mentionné dans la section précédente, les structures en raster 3D découpent la totalité de l'espace de modélisation en cellules. (Burrough et McDonnell 1998). À chaque cellule (voxel), sont attribués un ou plusieurs attributs dont la valeur est donnée par une méthode d'interpolation spatiale 3D. Le niveau de détail de la représentation des phénomènes géographiques sera donc amélioré à mesure que le nombre de voxels sur les axes X, Y, Z augmente. En ajoutant la troisième dimension à la structure de raster cependant, le nombre de cellules peut rapidement exploser et devenir un obstacle au temps de traitement et d'analyse de données. Néanmoins, le développement récent de la capacité des ordinateurs de bureau et des outils de modélisation géologique ont facilité l'accès à ce genre de représentation spatiale (Raiber et al. 2012; Wycisk et al. 2009).

Certains outils de modélisation de données géoscientifiques offrent la possibilité d'utiliser des voxels de forme irrégulière, ce qui permet d'améliorer la représentation d'un phénomène spatial (Falivene et al. 2007; Falivene, Cabrera, et Saez 2007). La Figure 6 présente deux modèles de raster 3D où l'espace de modélisation sur l'axe Z a été ajusté à la topographie du terrain. Dans la Figure 6a, le nombre de voxels en Z a été réduit (taille des voxels préservée). Dans la Figure 6b, les voxels ont été comprimés entre la surface et le fond, déformant leur taille (nombre de voxels préservé).

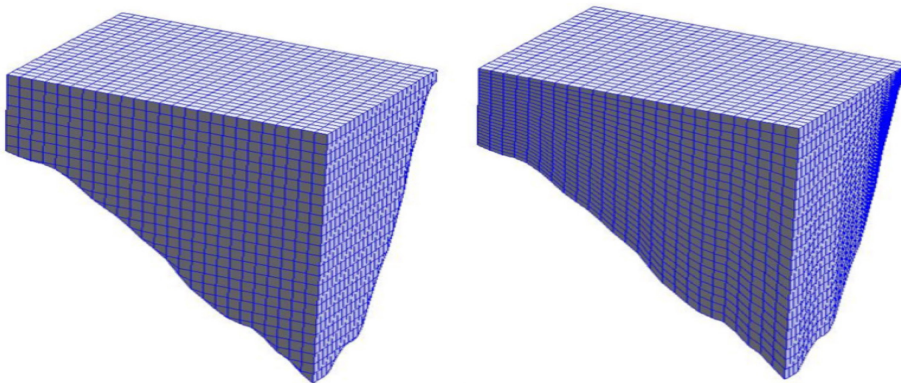


Figure 6. Modèles de données raster 3D avec voxels de forme régulière (*gauche*) et irrégulière (*droite*).

2.2.2.2 *Triangulation de Delaunay et Diagramme de Voronoï*

La triangulation de Delaunay (TD) et le diagramme de Voronoï (DV) constituent deux structures de données en tesselle alternatives à la structure raster, qui peuvent être adaptées au 3D (Hashemi Beni 2009; Ledoux et Gold 2006). Ces structures découpent l'espace en cellules irrégulières dont la forme est déterminée par la configuration spatiale des points d'échantillonnage (Hashemi Beni et al. 2011; Ledoux 2006; Mostafavi et Gold 2004). Les structures de TD et de DV sont ici présentées en 2D (Figure 7), mais les mêmes principes sont également valides en 3D (pour la TD, le triangle est remplacé par un tétraèdre, pour le DV, le polygone par un polyèdre). La TD est une méthode spécifique de triangulation basée sur la connexion des points dans l'espace en maximisant les angles petits (Ledoux 2006; Mostafavi et Gold 2004). Un triangle Delaunay est ainsi le plus *équilatéral* possible.

La structure de DV décompose l'espace dans des régions appelées cellules Voronoï à partir d'un ensemble de points. L'espace est subdivisé en des segments *équidistants* entre ces points (Hashemi Beni 2009; Ledoux 2006; Mostafavi et Gold 2004). Cela signifie que pour deux points voisins, un segment est créé exactement à mi-chemin entre les deux points perpendiculaires au segment formé par la frontière entre les deux points (Figure 7). Le résultat est une tessellation irrégulière formée de polygones (2D) où polyèdres (3D). Plusieurs auteurs ont évoqué les performances de la structure DV pour la gestion dynamique et temporelle des données spatiales marines (Gold et Condal 1995; Ledoux et Gold 2006; Mostafavi et Gold 2004). Il importe toutefois de mentionner que cette structure est encore peu implantée dans les SIG commerciaux et les outils de visualisation scientifique.

Le DV est étroitement lié à la TD (en 2D comme en 3D). La maximisation des petits angles dans la création de la TD rejoint le critère de la décomposition de l'espace par des segments équidistants entre les points du DV. Ainsi, chaque élément d'une cellule Voronoï est associé à un élément de triangle Delaunay. On dit ainsi qu'il existe une dualité entre le DV et la TD.

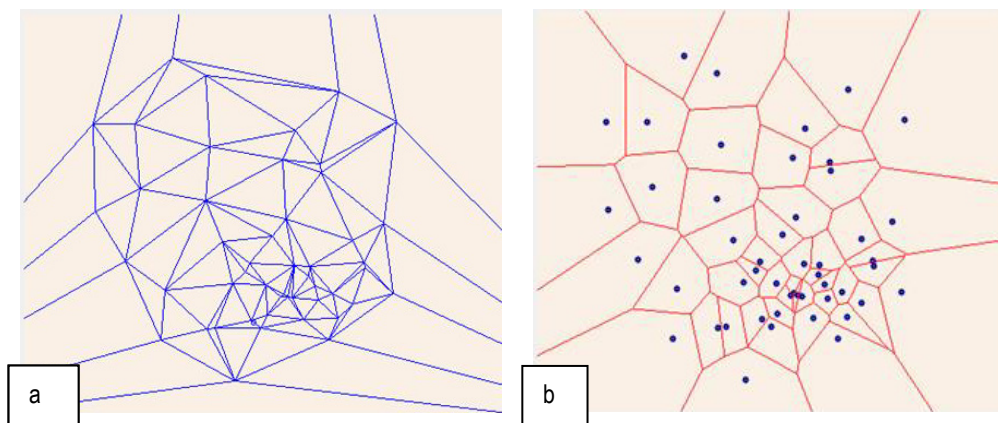


Figure 7. Triangulation de Delaunay (TD) et diagramme de Voronoï (DV). Le TD (a) et DV (b) pour un même ensemble de points en 2D. Dans a, les points (non illustrés) sont localisés aux vertex des triangles.

2.2.3 Outils d'analyse de modèles 3D de phénomènes continus

Les sections précédentes ont démontré que l'outil idéal de modélisation spatiale de l'écosystème pélagique marin prendrait en compte l'aspect 3D, continu et dynamique de cet environnement. Afin de répondre au premier sous-objectif de ce mémoire (*évaluer le potentiel et les limites des outils (logiciels) émergents pour la représentation, la visualisation et l'analyse 3D de l'écosystème pélagique marin*), une liste de critères d'évaluation de logiciel de modélisation spatiale 3D a été établi (Tableau 1). L'article 1 (chapitre 3) présente les résultats de l'inventaire, qui comprend des outils SIG, de visualisation scientifique et de modélisation géologique. Certains des critères ont déjà été traités en profondeur dans les sections précédentes alors que d'autres seront abordés dans cette section ou les sections subséquentes. Soulignons que l'aspect dynamique de l'écosystème pélagique n'a pas été utilisé directement comme un critère, la littérature actuelle révèle qu'aucun des logiciels de modélisation spatiale 3D examinés n'offre de telles fonctionnalités. Soulignons toutefois que des recherches sont actuellement menées dans ce sens, principalement avec la structure de données DV (Diagramme de Voronoï).

La pertinence des structures de données en tesselle (raster) pour la représentation spatiale des phénomènes continus a été abandonnée abordée dans les sections 2.2.1-2.2.2. Il a également été noté à la section 2.2.1 que la structure vectorielle peut parfois être une alternative ou un complément à la structure en tesselle pour la représentation 3D des phénomènes continus.

Tableau 1. Critères d'évaluation de logiciels de modélisation spatiale 3D de phénomènes continus.

Numéro	Critère
1	Représentation vectorielle 3D (structure de données vectorielle)
2	Représentation raster 3D (structure de données en tesselle)
3	Interpolation spatiale 3D
4	Analyse spatiale 3D
5	Analyse statistique (et géostatistique) 3D
6	Visualisation coupes
7	Visualisation <i>iso-surfaces</i>
8	Visualisation volumes

Tel que précisé précédemment, l'interpolation spatiale sert à reconstruire des champs continus à partir de données ponctuelles. L'interpolation spatiale est donc nécessaire pour des modèles de données en tesselle, comme le raster 3D. Elle est également un prérequis lorsqu'on veut utiliser une structure vectorielle comme les isosurfaces pour représenter un phénomène continu. La construction des champs 3D requière des méthodes d'interpolation adaptées à la troisième dimension. Compte tenu de l'emphase mise sur l'interpolation spatiale dans ce mémoire, une analyse exhaustive y sera consacrée dans la section 2.2.4.

On distingue trois méthodes principales pour la visualisation des champs continus 3D qui sont importants pour la représentation spatiale de l'écosystème pélagique, soit le découpage du volume par des plans 2D verticaux et/ ou horizontaux, la construction d'*isosurfaces* et la visualisation directe du volume par une image 2D. La visualisation des modèles spatiaux 3D sera spécifiquement abordée dans la section 2.2.5.

Les critères d'évaluation d'analyse spatiale 3D et d'analyse statistique 3D regroupent plusieurs type d'opérateurs. Cette approche s'explique par l'objectif de l'inventaire qui n'était pas d'énumérer en détails les opérateurs, mais de donner un aperçu général de la disponibilité des fonctionnalités 3D de base.

Il existe différentes façons de classier des opérateurs d'analyse spatiale dans la communauté de SIG (pour une revue de ces classifications, voir Beni 2009). Pour le modèle de données en tesselle, une des classifications correspond à l'algèbre de cartes, introduite par D. Tomlin vers la fin des années 1970 (Tomlin 1994). Dans la formalisation de l'algèbre de cartes, des opérations arithmétiques sur des cellules ou des groupes de cellules dans un raster sont différenciées par leur portée spatiale (p.ex. *fonctions locales* s'appliquant à une cellule, *fonctions de voisinage* s'appliquant à l'ensemble de cellules dans une couche,

etc.)¹⁰. Cette classification a été conçue pour le raster 2D, mais s'applique également au modèle de données 3D avec des modifications mineures (Ledoux et Gold 2006). Les opérateurs d'analyse spatiale de proximité, de génération de *buffer*, de classification et de pondération constituent tous des exemples pertinents à l'analyse du modèle 3D.

À l'instar des distinctions par portée spatiale, les analyses statistiques peuvent être classifiées. Celles-ci peuvent correspondre à des calculs de volumes (relativement simple pour une structure de raster 3D, mais complexe pour un modèle de données de triangulation 3D). On inclut également la géostatistique 3D dans cette catégorie, comprenant l'analyse variographique du comportement d'une variable spatiale. Cette dernière constitue une étape nécessaire à l'interpolation spatiale géostatistique, mais peut également être utilisée pour déduire de l'information spatiale d'une variable (voir section 2.2.4.4).

2.2.4 Interpolation spatiale

Pour un certain nombre de points d'échantillonnage avec un attribut a associé, l'interpolation spatiale est la procédure qui consiste à estimer la valeur de l'attribut à une localisation spatiale non-échantillonnée par la création d'une fonction f qui relie ces points ensemble (Ledoux 2006). Par conséquent, l'objectif de l'interpolation spatiale peut être décrit ainsi: *Given a set of spatial data either in the form of discrete points or for subareas¹¹, find the function that will best represent the whole surface and that will predict values at other points or for other subareas* (Lam 1983).

Dans cette section, une attention particulière est portée aux méthodes d'interpolation *géostatistiques*, en raison de (1) leur place importante dans le domaine de l'interpolation spatiale et (2) leur complexité comparée à des méthodes d'interpolation de distance inversée (IDW). Soulignons qu'aucune étude scientifique portant spécifiquement sur l'interpolation spatiale 3D de l'écosystème pélagique marin n'a été recensée au moment de cette recherche. Cette section reflète cette réalité. Finalement, des techniques de validation de la performance (la minimisation des erreurs de l'interpolation et la stabilité de la méthode par rapport au paramétrage) des méthodes d'interpolation sont présentées au chapitre 4.

Les méthodes d'interpolation spatiale connaissent présentement un développement rapide (MacCormack et Eyles 2012) et leur utilisation est commune dans toutes les sciences de la terre et environnementales (Li et Heap 2011; Zhou et al. 2007). Le choix de la méthode et de paramètres spécifiques influence toutefois les

¹⁰ Pour plus de détails sur cette classification, voir Tomelin (1994).

¹¹ Lam fait une distinction entre l'interpolation spatiale de points (*point interpolation*) et de données de surface (*areal data*). À travers ce mémoire, l'*interpolation spatiale* fait référence uniquement à l'interpolation spatiale de points.

résultats de l'interpolation spatiale et donc la précision de la modélisation (Gotway et al. 1996; Li et Heap 2008, 2011; Reed, Ellsworth, et Minsker 2004). Il a été démontré récemment que même le choix du logiciel utilisé pour l'interpolation peut avoir un impact sur les performances (MacCormack et Eyles 2012). De plus, le nombre de méthodes d'interpolation disponible est en outre très important : Li et Heap (2011) ont identifié 72 méthodes/sous-méthodes d'interpolation spatiale utilisées dans les sciences environnementales pour l'étude des variables continues. C'est pourquoi le résultat de la modélisation spatiale repose largement sur l'expérience de l'utilisateur.

Alors que le nombre de publications traitant des performances des méthodes d'interpolation dans l'espace 2D est relativement élevé, il demeure faible en 3D. Cela peut s'expliquer par les limites des outils qui jusqu'à récemment étaient restreints à l'environnement 2D : les produits commerciaux de SIG ont été créés à l'origine pour représenter l'environnement terrestre en 2D dans un environnement statique (Arsenault et al. 2004; Basu 1996; Carette, Mostafavi, et Devillers 2008; Gold et Condal 1995; Mesick et al. 2009; Ng et al. 2007; Wright et Goodchild 1997).

2.2.4.1 Classification des méthodes d'interpolation

Selon la méthode de classification la plus commune, on distingue deux approches principales d'interpolation spatiale, soit (1) les méthodes déterministes ou non géostatistiques et (2) les méthodes stochastiques ou géostatistiques. Les méthodes déterministes sont basées sur des fonctions mathématiques et sont appréciées pour leur implantation simple et directe (Murphy, Curriero, et Ball 2010). Les méthodes de TIN (Triangular Irregular Network), de distance inverse (IDW), de voisin plus proche et de voisin naturel représentent quatre méthodes déterministes répandues. Les méthodes géostatistiques combinent des fonctions mathématiques et des analyses statistiques; elles seront abordées à la section 2.2.4.2. Quelques méthodes d'interpolation qui combinent des méthodes déterministes et des méthodes géostatistiques (p.ex. *Regression Kriging*) pourraient constituer un troisième groupe de méthodes d'interpolation (Li et Heap 2011). Ce groupe ne sera toutefois pas traité dans ce mémoire.

Les estimations des méthodes d'interpolation spatiales mentionnées ci-haut peuvent être exprimées comme une moyenne pondérée des points d'échantillons (Ledoux 2006; Li et Heap 2008).

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i(x_0)z(x_i) \quad (1)$$

, où \hat{z} représente la valeur estimée à la localisation (x_0), λ_i le poids attribué à la localisation (x_0), pour le point de mesure (x_i), $z(x_i)$ la valeur de la mesure à la localisation (x_i) et n le nombre de points voisins participants dans l'estimation.

Les méthodes d'interpolation peuvent également être classifiées selon leurs caractéristiques générales. On les qualifie de *globales* lorsque l'ensemble des points de mesure de la région d'étude est utilisé dans l'estimation à une localisation non échantillonnée, de *locales* lorsqu'un nombre limité de points de mesure est utilisé pour l'estimation (voisinage local, Li et Heap; Lam 1983). Une méthode d'interpolation est considérée *exacte* lorsque l'estimation reproduit les valeurs mesurées à un point de mesure, et *inexacte* (ou *approximative*) dans le cas contraire (Lam 1983; Li et Heap 2008). La méthode d'interpolation peut également être *univariée*, lorsque la variable principale est utilisée dans l'estimation, ou *multivariée*, lorsque des variables secondaires sont incluses.

2.2.4.2 La géostatistique

La *géostatistique* est née dans les années 1950 à la suite des travaux de D.G. Krige, H.S. Sichel et plus tard de G. Matheron. Ce domaine de recherche, qui constitue une branche de la statistique, peut être défini comme *l'ensemble de techniques statistiques d'estimation et de simulation s'appliquant à des champs variables dans l'espace et/ou dans le temps* (OQLF 2013).

Si la géostatistique a été développée à l'origine spécifiquement pour la modélisation spatiale du sous-sol géologique, elle est aujourd'hui largement utilisée à travers l'ensemble des sciences naturelles. Selon Zhou, Guo et al. (2007), les dix domaines de recherche les plus fréquemment rencontrés dans le domaine de la géostatistique (ordre d'importance décroissante) sont les suivants : (1) les sciences de la terre (multidisciplinaires), (2) les ressources hydrauliques, (3) les sciences environnementales, (4) les sciences du sol et de l'agriculture, (5) les mathématiques et des applications interdisciplinaires, (6) la statistique et la probabilité, (7) l'écologie, (8) le génie civile, (9) le génie minier et (10) la limnologie.

Comme mentionné auparavant, les méthodes géostatistiques combinent des fonctions mathématiques et des analyses statistiques. On cherche ainsi une fonction numérique (déduite de la modélisation du *semi-variogramme* ou *variogramme*¹², section suivante) pour expliquer la variabilité spatiale locale du phénomène. En intégrant l'information sur le comportement de la variable dans le processus d'interpolation on peut ainsi améliorer l'estimation dans certains cas. La grande différence entre les méthodes géostatistiques et les méthodes déterministes réside donc dans la distribution du poids des points voisins utilisés dans l'estimation : elle est basée non seulement sur la distance, mais également sur la localisation géographique des mesures. Ainsi, les méthodes géostatistiques prennent en considération la configuration spatiale des points d'échantillonnage, ce qui n'est pas le cas pour les méthodes déterministes. L'estimation d'une mesure d'erreur

¹² Pour simplifier, *variogramme* est souvent utilisé à la place du terme précis *semi-variogramme* et serait. Selon certains auteurs, le terme *variogramme* serait également préférable (Bachmaier et Backes 2008).

pour chaque point interpolé – la *variance de krigeage* – constitue un autre avantage des méthodes géostatistiques (Murphy, Curriero, et Ball 2010; Reed, Ellsworth, et Minsker 2004).

Les méthodes d'interpolation géostatistiques regroupent un nombre important de méthodes de *krigeage* et des nouvelles variantes sont continuellement décrites. Parmi les méthodes les plus fréquemment utilisées, on note le krigeage ordinaire, le krigeage universel et le cokrigeage ordinaire (Li et Heap 2011; MacCormack, Brodeur, et Eyles 2013). Pour l'ensemble de ces méthodes, l'objectif est d'estimer le plus précisément possible la valeur de l'attribut recherché par une moyenne pondérée des observations du voisinage¹³, selon des variantes de l'équation (1). Dans ce mémoire, seulement le krigeage ordinaire sera abordé plus en détail (section 2.2.4.5).

L'interpolation par krigeage peut être divisée en deux étapes distinctes, soit (1) la construction du variogramme et l'analyse variographique et (2) le processus d'estimation (interpolation). Ces deux étapes seront abordées dans les prochaines sections.

2.2.4.3 Modélisation du variogramme 2D

L'étude de la structure spatiale des variables par le calcul du variogramme (analyse variographique) est cruciale, car elle permet postérieurement d'estimer les valeurs inconnues de la variable régionalisée¹⁴ à utiliser pour l'estimation des poids de krigeage. D'autres implications directes (qui ne nécessitent pas l'interpolation) comprennent l'identification de l'anisotropie géométrique et zonale et l'estimation de l'autocorrélation spatiale (voir Gringarten et Deutsch 2001 pour une bonne revue de ces concepts). Un exemple d'utilisation de ces informations correspond à l'adaptation d'un plan d'échantillonnage par l'augmentation de la densité des points dans l'azimut où l'autocorrélation est le plus faible.

Différents types de prétraitements des jeux de données permettent également de les rendre plus adéquats aux méthodes d'interpolation géostatistiques. Ainsi, on utilise souvent une méthode de normalisation de données pour transformer le jeu de données afin qu'il ressemble à une distribution normale standard (Gringarten et Deutsch 2001). Cette transformation apporte deux avantages majeurs dans l'analyse du variogramme : 1) la différence entre les valeurs extrêmes diminue et 2) le palier du variogramme devient standardisé à une valeur de « 1 » (Figure 8; Gringarten et Deutsch 2001). Il existe différentes méthodes de normalisation de données (voir p.ex. (Daumas 1982)). Par conséquent, la disponibilité et le choix des

¹³ Le voisinage, dans le contexte de l'interpolation spatiale, fait référence au champ qui contient le site à estimer et les points utilisées pour l'interpolation.

¹⁴ On définit une *variable régionalisée* comme étant une fonction numérique prenant ses valeurs dans une région limitée (champ).

méthodes varient d'un outil de modélisation géostatistique à un autre. GOCAD® (version 2009.2) utilise la méthode de standardisation de données par l'écart réduit, décrit par l'équation suivante:

$$Y = \frac{Z - m}{\sigma} \quad (2)$$

où Y représente la valeur normalisée du point de mesure Z tandis que m et σ représentent la moyenne et l'écart-type de la variable respectivement.

La variance (ou semi-variance¹⁵, γ) entre deux points de mesure séparée d'un vecteur de distance h exprime le degré de dépendance entre les points. Le variogramme d'une variable régionalisée décrit la variation de la variance en fonction de h et est définie par l'équation :

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n (z(x_i) - (x_i + h))^2 \quad (3)$$

où n correspond au nombre de paires séparés par l'intervalle de distance (p.ex. 100 à 200m, 200 à 300m, etc.), (x_i) , $(x_i + h)$ représentent les points de mesure et $\gamma(h)$ représente le variogramme expérimental (ou variogramme *empirique*) (Burrough et McDonnell 1998; Deutsch 2002; Isaaks et Srivastava 1988; Lam 1983).

Lorsque le variogramme expérimental est construit, on y ajuste une fonction mathématique (le *modèle* du variogramme) soumise à certaines contraintes¹⁶. Cette étape se nomme la *modélisation* du variogramme. Le modèle peut ainsi donner un indice sur le type de variation spatiale présente dans une région donnée. Parmi les fonctions les plus communes, on note les modèles sphérique, exponentiel, gaussien et linéaire. La modélisation du variogramme est nécessaire afin de déduire *la valeur de la variance*, utilisée pour le calcul des poids de l'équation du krigeage. Trois paramètres sont déduits du variogramme, soit la *portée*, le *palier* et la *pépite*. Le *palier* (C_1) correspond à la dissemblance maximale du modèle (la valeur de γ maximale). La *portée* (a) se définit par la distance à partir de laquelle le variogramme atteint son palier. À cette distance, deux valeurs ne sont plus corrélées et par conséquent, un échantillon sur une distance excédant la portée n'apporte plus de contribution à l'estimation. Cela est indiqué dans la Figure 8 par un coefficient de corrélation dans les diagrammes, qui diminue avec une augmentation de la valeur h . La *pépite* (C_0) correspond à la valeur positive du modèle lorsqu'il traverse l'axe Y . Une valeur de $C_0 > 0$ implique que même à des points d'échantillonnage séparés par une très faible distance, les valeurs diffèrent. Cela peut s'expliquer par une

¹⁵ Le terme *variance* est à préférer selon certains auteurs (voir Bachmaier et Backes (2008) pour une discussion approfondie sur la terminologie) et sera utilisé dans ce mémoire.

¹⁶ La fonction doit, entre autres, être une fonction de type *négalif conditionnelle*. Pour plus d'informations au sujet de ces contraintes, voir Desassis et Renard (2012).

variabilité naturelle de la variable (p.ex. deux valeurs différentes sont mesurées à deux moments différents) ou par l'erreur de la mesure (erreur de l'instrument, ou de sa manipulation, etc.).

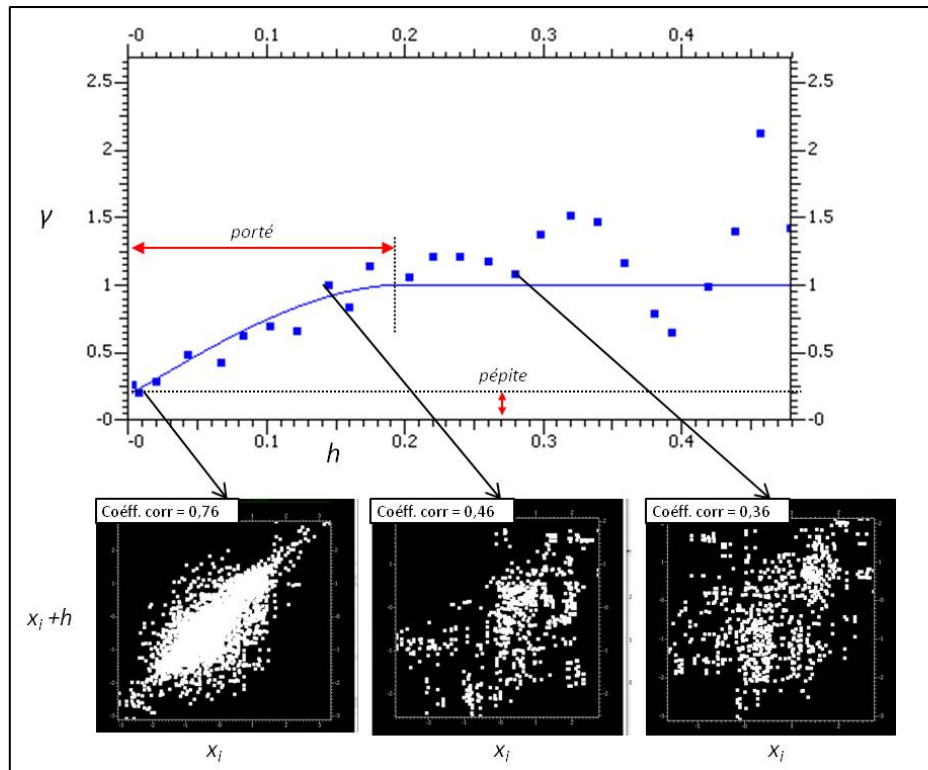


Figure 8. Le variogramme. Exemple de variogramme avec 3 diagrammes de dispersion d' h pour trois intervalles d' h (*lags*) distincts. Les points correspondent au variogramme *expérimental*, alors que la ligne entière représente l'ajustement du variogramme (modèle sphérique). La variable (c_p) a été transformée en une distribution gaussienne normale.

Les paramètres C_0 , C_1 et a sont utilisés pour calculer la valeur du variogramme pour h , tel qu'illustré ici avec la fonction sphérique.

$$\gamma(h) = C_0 + C_1 \left\{ \frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right\} \quad \text{pour } 0 < h < a$$

$$\gamma(h) = C_0 + C_1 \quad \text{pour } h \geq a \quad (4)$$

$$\gamma(0) = 0$$

Pour une localisation x_0 pour laquelle on souhaite une estimation avec la méthode de krigeage, la fonction mathématique est donc utilisée pour déduire la variance qui sert à calculer le poids à attribuer à chaque point de mesure voisin $z(x_i)$ participant dans l'estimation.

L'ajustement du modèle mathématique au variogramme expérimental et le choix du modèle sont parmi les éléments les plus controversés en géostatistique (Webster et Oliver 2007) et constituent un sujet de recherche actuel (Dahle, Kolbjornsen, et Abrahamsen 2005; Desassis et Renard 2012; Miranda et de Miranda 2011). Un ajustement par la méthode des moindres carrés est la méthode classique (Burrough et McDonnell 1998; Cressie 1985; Miranda et de Miranda 2011). Une critique de cette méthode stipule qu'on devrait accorder plus de poids aux h qui correspondent à un nombre plus élevé de paires. En ce sens, des méthodes d'ajustement de moindre carrée *pondérées* et *généralisées* ont été développées (Cressie 1985; Lahiri, Lee, et Cressie 2002). Miranda et de Miranda (2011) soulèvent des problèmes de stabilité avec ces méthodes et proposent plutôt une modification de la méthode de Cressie. Webster et Oliver (2007) constatent qu'une stratégie d'ajustement *by eye* est encore commune et suggèrent qu'un ajustement statistique (i.e. moindre carrée) devrait être validé par un examen visuel. Dahle et al. (2005) rapportent qu'un ajustement automatique des variogrammes 3D peut parfois mener vers un modèle de variogramme qui est non conforme à la réalité. Cela serait surtout dû au nombre important de données requis pour estimer certains des paramètres du variogramme 3D (notamment l'angle de rotation verticale, voir section suivante). Il est à noter que dans le logiciel de modélisation GOCAD® (ver. 2009.2) utilisé dans le cadre de cette recherche, aucun ajustement automatique des variogrammes 3D n'est disponible et par conséquent, cet ajustement s'effectue *by eye*.

Dans l'étude spatiale du milieu marin (et d'autres environnements naturels), on observe fréquemment une variation de la portée dans différents azimuts (Gringarten et Deutsch 2001; Jensen, Christman, et Miller 2006). On qualifie ce phénomène d'*anisotropie géométrique* (*isotropique* dans le cas contraire). L'anisotropie géométrique est étudiée en utilisant un variogramme directionnel, soit un variogramme modélisé dans différents intervalles d'azimuts séparés par une rotation successive de l'azimut central (Figure 9a). Pour chaque intervalle d'azimut, un variogramme expérimental est obtenu. Ces variogrammes serviront à identifier un axe X de portée maximale et un axe Y perpendiculaire de portée minimale. Le résultat est souvent illustré par une ellipse de la dépendance spatiale de la variable avec un axe majeur (portée maximale), un axe mineur (portée minimale; Figure 9b). Lors de la modélisation d'un variogramme 3D, cette ellipse devient un ellipsoïde en ajoutant un axe Z de portée verticale.

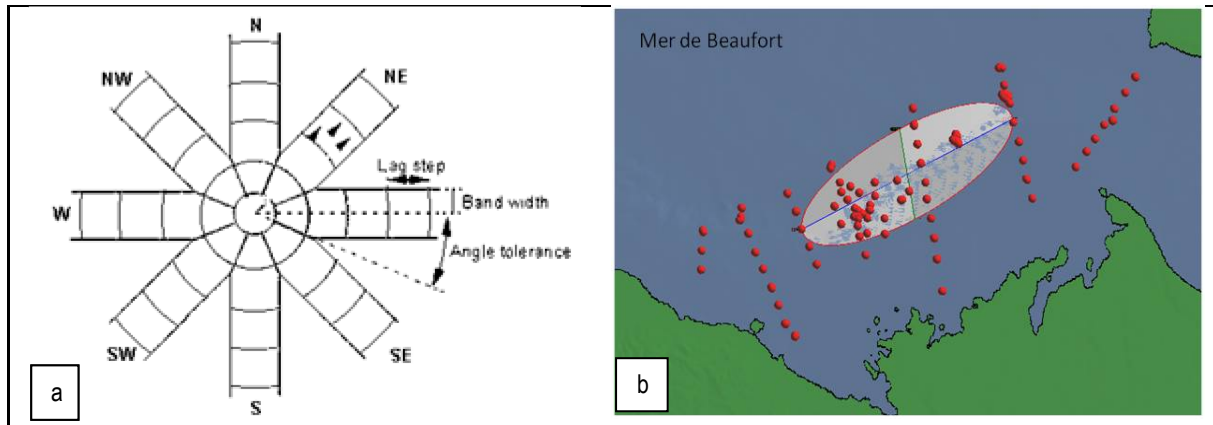


Figure 9. Variogramme directionnel, cas 2D. (a) Nomenclature du variogramme directionnel où lag step correspond à un intervalle d' h ; Band width à la largeur du cône de tolérance. (b) Ellipse illustrant les portées de dépendance spatiale déduites d'un variogramme directionnel 2D.

2.2.4.4 Modélisation du variogramme 3D

La modélisation du variogramme 3D s'effectue sur les mêmes principes que le variogramme 2D, mais la complexité additionnelle a été soulignée par plusieurs auteurs (Dahle, Kolbjornsen, et Abrahamsen 2005; Desassis et Renard 2012; Mucha et Wasilewska-Blaszczyk 2012). Les approches résumées ici sont celles proposées par le logiciel Paradigm GOCAD® (ver. 2009.2), bien que d'autres approches puissent exister.

On note deux approches distinctes de la construction du variogramme 3D. Dans le premier cas, on modélise le variogramme vertical indépendamment du variogramme horizontal (aussi appelé variogramme 1D+2D). Pour l'étape de la construction du variogramme vertical (1D), on divise l'espace en colonnes, sur les axes X et Y (p.ex. 100 x 100 colonnes), de sorte que dans chaque cellule se situe au maximum une seule station d'échantillonnage (une station d'échantillonnage comprenant un à plusieurs points de mesure à la verticale). La variance à la verticale est calculée individuellement sur chaque colonne entre les paires de points de mesure sur les intervalles de distance h . Les différentes valeurs sont par la suite combinées dans un variogramme 1D. La partie du variogramme 2D se construit par une stratification du modèle volumétrique par des couches spécifiées par l'utilisateur (p.ex. 100 strates, Figure 10). Comme pour le variogramme vertical, la variance ($\gamma(h)$) est calculée individuellement pour chaque strate, puis combinée dans un variogramme 2D. Ce variogramme peut être isotrope ou anisotrope. La stratification est effectuée à partir de l'espace correspondant le mieux à la dispersion de la variable à l'étude dans l'espace selon nos connaissances du milieu. En prenant comme exemple la température de l'écosystème pélagique, on utiliserait une stratification par couches horizontales, puisqu'on peut normalement présumer que la température sera principalement stratifiée sur la verticale en couches parallèles à la surface de l'eau. L'approche de 1D+2D a été utilisée par plusieurs auteurs (Falivene et al. 2007; He et al. 2009; Mucha et Wasilewska-Blaszczyk 2012), mais il ne permet pas de déduire dans quelle angle de rotation verticale par rapport au plan horizontal (*dip angle*) la

portée est la plus importante (Gringarten et Deutsch 2001), ce qui est possible avec la deuxième approche de modélisation. Il a toutefois été souligné que le nombre de points de mesure dans l'espace doit être grand afin que ce paramètre puisse être estimé adéquatement (Dahle, Kolbjornsen, et Abrahamsen 2005).

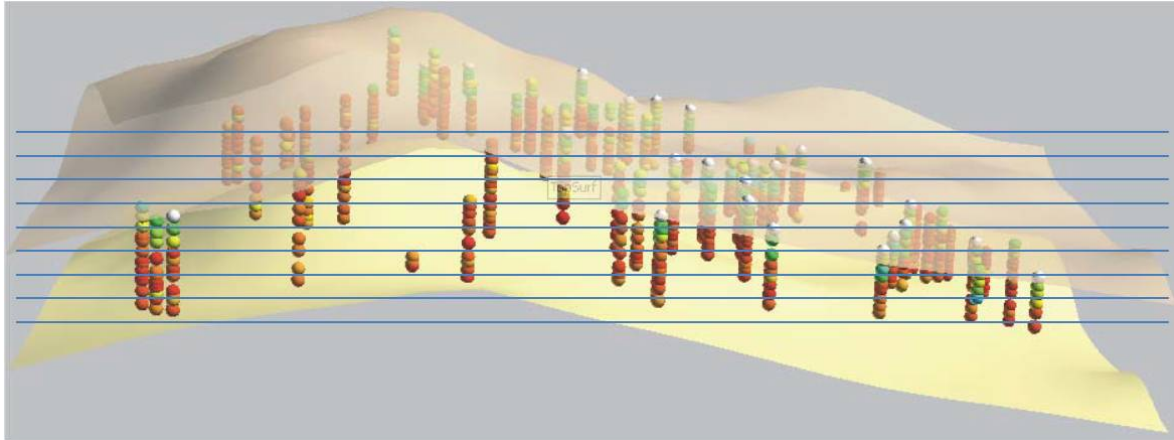


Figure 10. Stratification de l'espace de modélisation pour l'estimation du variogramme 1D+2D. (Bergacademie XXXX).

Dans la deuxième approche (variogramme 3D), on crée le variogramme dans différents angles de rotation verticale par rapport au plan horizontal. Le principe est le même que pour le variogramme directionnel (Figure 9), sauf que l'intervalle de modélisation autour d'un azimut central (dans le variogramme directionnel) est remplacé par un intervalle de modélisation autour d'un angle de rotation verticale (p.ex. 0, 45 et 85° : valeurs par défaut dans GOCAD®). Ainsi, lorsque le variogramme 3D comprend un variogramme anisotrope dans le plan horizontal, le nombre de variogrammes expérimentaux se calcule par la multiplication du nombre d'intervalles de modélisation horizontaux par le nombre d'angles de rotation verticale (pour une modélisation dans 4 azimuts et 3 angles de rotation verticale, on obtiendra donc 12 variogrammes expérimentaux, Figure 11).

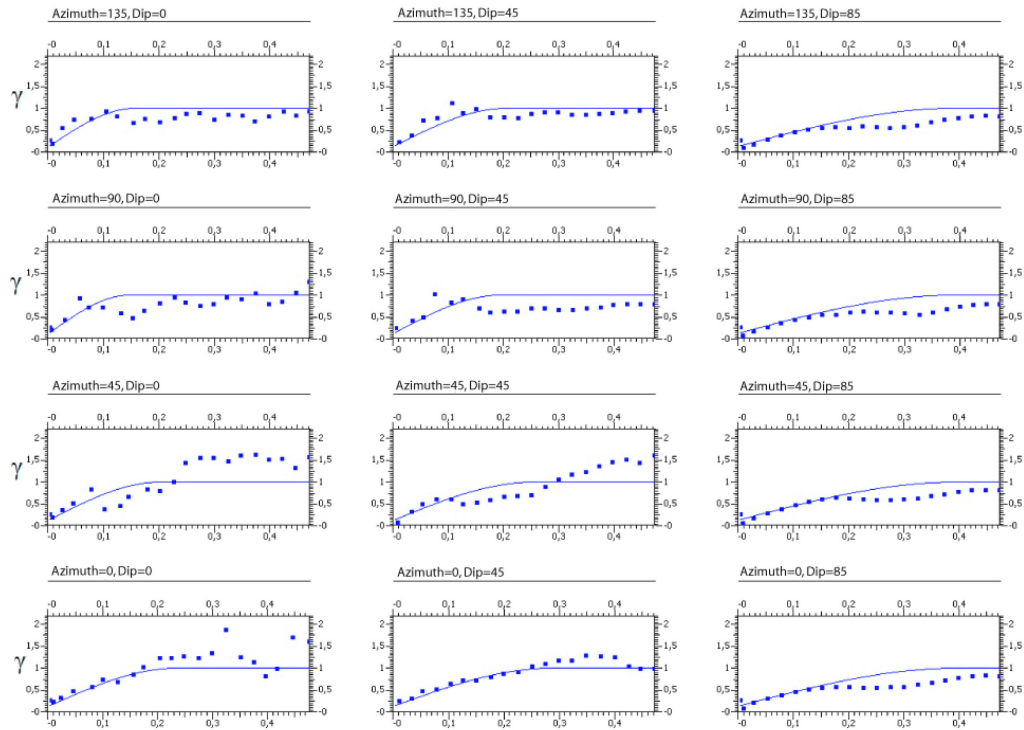


Figure 11. Variogramme 3D. Variogramme modélisé dans quatre intervalles d'azimuts (azimut central : 0 °, 45 °, 90 °, 135°) et trois intervalles d'angles de rotation verticale (0 °, 45 °, 85 °). Modèle du variogramme = sphérique. La variable (c_p) a été transformée en une distribution gaussien normalisée L'axe Y correspond au γ . L'axe X représente la fraction de l'espace de modélisation (*grid*) utilisée pour la modélisation du variogramme.

Pour l'ajustement de la fonction mathématique au variogramme 3D, on note que : (1) la même fonction mathématique est utilisée pour l'ensemble des variogrammes expérimentaux, (2) la pépité devrait être estimée dans l'angle de rotation dans lequel on a le plus de points de mesure (habituellement dans le sens vertical), (3) le pallier est toujours égale à « 1 » si la variable analysée est normalisée et (4) la pépité prend la même valeur pour l'ensemble des variogrammes expérimentaux (Dahle, Kolbjornsen, et Abrahamsen 2005; Gringarten et Deutsch 2001). La fonction mathématique qui s'ajuste le mieux aux variogrammes expérimentaux est choisie.

Dans les deux approches (variogramme 2D+1D ou 3D), les paramètres déduits de l'analyse variographique sont utilisés dans la fonction mathématique (p.ex. équation 4) pour calculer le poids à attribuer à chaque point voisin participant dans le processus d'estimation d'un point inconnu. Les paramètres suivants sont retenus de l'analyse variographique :

1. le modèle du variogramme (sphérique, etc.);

2. la portée 1 (axe majeur de l'ellipsoïde dans le plan horizontal);
3. la portée 2 (axe mineur de l'ellipsoïde dans le plan horizontal);
4. la portée 3 (axe vertical de l'ellipsoïde);
5. l'azimut de l'axe majeur de l'ellipsoïde;
6. l'angle de rotation verticale avec la portée maximale (seulement variogramme 3D).

Si la variable est isotropique dans le plan horizontal (et alors, portée1 = portée2), l'azimut (paramètre 2) ne sera pas utilisé par les équations de l'interpolation de krigeage. Si le variogramme 3D est dérivé d'une modélisation 1D + 2D, l'angle de rotation verticale (paramètre 6) sera toujours zéro. Une valeur de zéro pour ce paramètre indique que la plus grande portée se trouve dans le plan horizontal. Alors qu'une seule valeur de portée est retenue du variogramme isotropique et deux du variogramme directionnel horizontal¹⁷, on retient trois portées pour le variogramme anisotropique 3D (ainsi que deux angles de rotation de l'ellipsoïde). Par des calculs trigonométriques, on combine les portées pour un point d'estimation x_i à l'aide des angles qui positionnent l'ellipsoïde dans l'espace. Cette opération doit donc être fait pour chaque point de mesure x_i participant dans une estimation (la distance et angle entre x_i et x_0 sont calculés par rapport à la rotation de l'ellipsoïde ce qui permet de calculer une portée pondérée pour x_i).

Les incertitudes dans l'estimation du variogramme 3D résultent de plusieurs facteurs, dont la qualité des mesures, la densité et la répartition des points de mesure dans la région d'étude, la variance des variables et le choix de modèle mathématique. Typiquement, pour un jeu de données issu d'une campagne océanographique, la répartition des points de mesure dans l'espace constitue un facteur majeur pour la qualité du variogramme : la densité de données à la verticale est de l'échelle de quelques mètres (ou dizaines de mètres), tandis que la distance à l'horizontal est de l'ordre de dizaines de kilomètres. La distribution critique est donc généralement le nombre de stations d'échantillonnage à l'horizontal. Comme le concluent Dahle, Kolbjornson et al. (2005), l'estimation de variogrammes en 3D est "étonnamment difficile". Pour un jeu de données générées à partir de variogrammes connus, ces auteurs démontrent que même lorsque 200 stations horizontales d'échantillonnage sont utilisées (densité élevée à la verticale), l'incertitude totale de la modélisation du variogramme est relativement importante. Alors qu'environ 50 stations seraient suffisantes pour obtenir des valeurs fiables des portées horizontale et verticale, environ 200 stations sont nécessaires pour obtenir une estimation crédible de l'angle de rotation verticale.

¹⁷ Le nombre de portées déduites du variogramme directionnel dépend du nombre de directions dans lesquelles le variogramme a été calculé, mais seules les portées maximale et minimale (et les azimuts correspondants) sont utilisées lors de l'interpolation avec OK. En 3D, une portée de l'axe Z s'ajoute.

2.2.4.5 Équation de Krigeage

Le krigeage présume une *autocorrélation spatiale* et une *stationnarité* des variables spatiales. L'*autocorrélation spatiale* implique la présomption que des points localisés plus près l'un de l'autre ont des valeurs plus rapprochées que des points plus distants (Goodchild 1992; Tobler 1970). Cela constitue la base du variogramme, et même les statistiques de l'environnement de façon générale. La condition de *stationnarité* est rencontrée pour le krigeage ordinaire lorsque la moyenne et la variance sont spatialement constantes à travers le voisinage¹⁸. On présume également une variable de moyenne inconnue.

Tel que mentionné, l'équation (1) constitue l'équation générale des méthodes d'interpolation. La différence entre les méthodes se trouve ainsi au niveau de la définition des poids (qui, pour les méthodes géostatistiques, nécessite les résultats de l'analyse du variogramme) et des conditions entourant l'équation. Les poids de krigeage ordinaire sont estimés avec deux conditions de base (Lam 1983), soit (1) la contrainte d'optimalité, ce qui réfère à la minimisation de l'erreur de l'estimation (la variance) et (2) la contrainte d'absence de biais (la somme des poids est égale à 1 (équation 5).

$$\sum_{i=1}^n \lambda_i = 1 \quad (5)$$

La contrainte d'optimalité requiert la recherche des extrema du lagrangien, ce qui nous amène à l'équation (6) :

$$\sum_{j=1}^n \lambda_j \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0) \quad \text{pour } i = 1, 2, \dots, n \quad (6)$$

où λ_i représente le poids de krigeage, $\gamma(x_i, x_j)$ la variance entre les points mesurés x_i et x_j , $\gamma(x_i, x_0)$ représente la variance entre le point de mesure x_i et le point non mesuré x_0 , μ représente le multiplicateur de Lagrange (μ) et n le nombre de points dans le voisinage.

La variance σ^2 de l'estimation à la localisation (x_0) est donnée par l'équation 7 :

$$\sigma^2(x_0) = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \mu \quad (7)$$

¹⁸ Le voisinage d'une méthode d'interpolation de krigeage fait référence au champ qui contient le site à estimer et les points utilisés pour l'interpolation.

La variance de l'estimation dépend donc des paramètres du krigeage (poids), ainsi que de la structure spatiale de la fonction mathématique (variogramme).

Les dérivations des équations 6 et 7 ne seront pas présentées dans ce mémoire en raison de leurs longueurs et des explications fastidieuses. Par conséquent le lecteur est référé à des ouvrages de référence (Cressie et Zimmerman 1992; Isaaks et Srivastava 1992; Kelkar et Perez 2002).

En pratique, la condition de *stationnarité* est souvent infirmée, par exemple en présence d'une anisotropie géométrique ou d'une stratification des variables (Jensen, Christman, et Miller 2006). Toutefois, l'adoption de correctifs pour remédier à ce problème constitue maintenant un processus inhérent à la modélisation géostatistique (*idem.*). Parmi ces correctifs, on peut mentionner l'utilisation du variogramme directionnel et la normalisation de données discutées précédemment. Notons également que le critère de stationnarité est plus souple pour le krigeage universel alors qu'il permet de traiter une variable non-stationnaire avec une moyenne à la *dérive* (tendance spatiale).

2.2.5 Visualisation de volumes

La transition d'une modélisation 2D vers une modélisation 3D de l'écosystème pélagique marin permet une représentation plus réaliste des variables continues. Cette transition implique toutefois des contraintes particulières en ce qui concerne le domaine de visualisation volumétrique (Kaufman 1996; Ledoux et Gold 2008). Ces contraintes sont à l'égard de la complexité des algorithmes de visualisation, mais également sur le plan du temps d'affichage et de traitement (Kaufman 1996).

Puisqu'aucune vue 2D isolée d'un espace volumétrique ne permet de déceler l'ensemble des relations spatiales à l'intérieur d'un volume, le recours à des techniques interactives pour manipuler les modèles spatiaux 3D est primordial (Kaufman 1996). Aujourd'hui, ce genre de fonctionnalités (translation, rotation et zoom) est toutefois intégré aux outils de modélisation spatiale.

En plus de ces fonctionnalités interactives, on distingue trois méthodes principales pour la visualisation des champs continus 3D, soit (1) la visualisation directe du volume par une image 2D (Figure 12); (2) la construction d'*isosurfaces* (Figure 13); (3) le découpage du volume par des plans 2D verticaux et/ ou horizontaux (Figure 14).

Tel que décrit par Ledoux et Gold (2008), la première méthode consiste à désigner dans le volume pour chaque voxel une couleur et une opacité (selon la valeur de l'attribut du voxel). La couleur d'un pixel spécifique de l'image est obtenue en additionnant les attributs de couleur qui sont projetés dans un même pixel. L'*isosurface* est l'équivalent 3D d'une *isoligne* sur une carte 2D et correspond donc à la surface

composée des voxels où l'attribut a une valeur constante (Ledoux et Gold 2006). La méthode par découpage du volume par des plans 2D verticaux et/ ou horizontaux donne un résultat qui ressemble à la représentation habituelle des variables de l'écosystème pélagique marin (section 1.1). Notons toutefois une différence importante : lors du découpage d'un volume, l'utilisateur peut choisir le nombre, l'angle et la dimension des coupes, ce qui apporte une continuité dans l'espace. Cette dernière est impossible lorsque la coupe est dérivée d'un jeu de données 2D (p.ex. le long d'un transect d'échantillonnage spécifique).

À ces méthodes de visualisation, je propose également d'ajouter à la classification des approches *le découpage du volume par des surfaces 2,5D* (surface courbée). Dans le cas de la visualisation de l'écosystème pélagique marin, il peut notamment être pertinent de visualiser la variabilité d'un phénomène dans la couche d'eau située le plus près du fond (Figure 14) ou le long d'un phénomène spécifique, comme la limite entre deux masses d'eau. Dans le cas de la visualisation des variables de l'écosystème pélagique marin, il est également primordial de recourir à un outil qui permet de combiner ces méthodes de visualisation volumétrique avec des outils sophistiqués de représentation, d'analyse et de visualisation des images satellites. En conclusion, la combinaison des techniques interactives et des différentes méthodes de visualisation volumétrique et 2D permettra de maximiser le résultat de l'analyse visuelle du modèle spatial 3D.

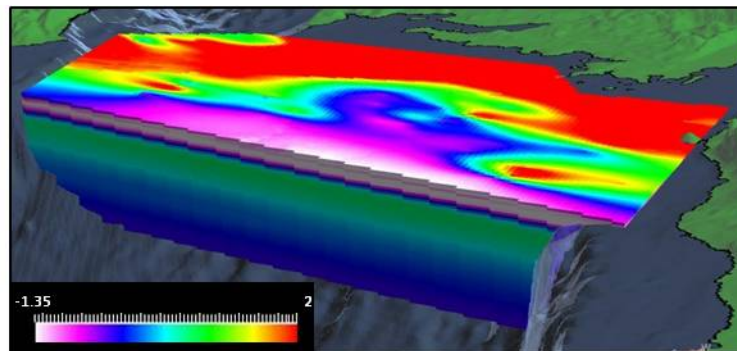


Figure 12. Visualisation d'un volume par image 2D (température de l'eau, °C).

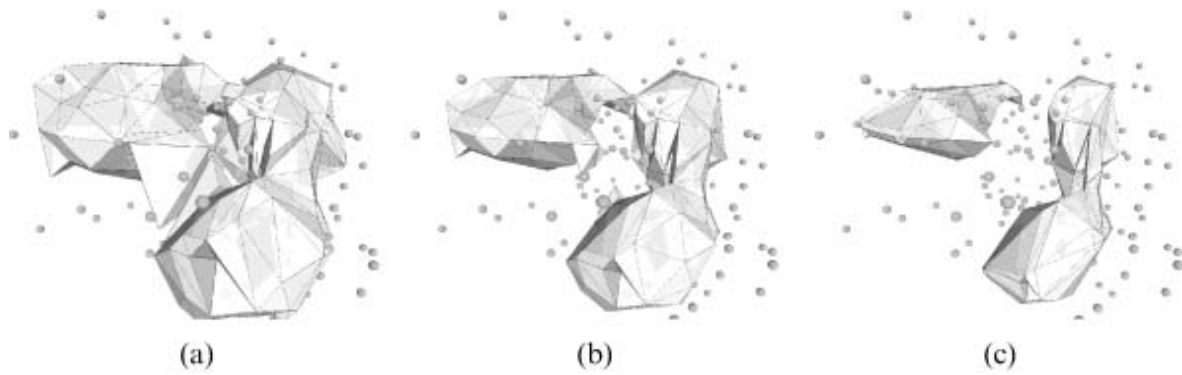


Figure 13. Exemple d'isosurfaces pour trois températures de l'eau différents. Températures : 7, 8 et 9°C pour (a), (b) et (c). La Figure a été tirée de Ledoux et Gold 2008.

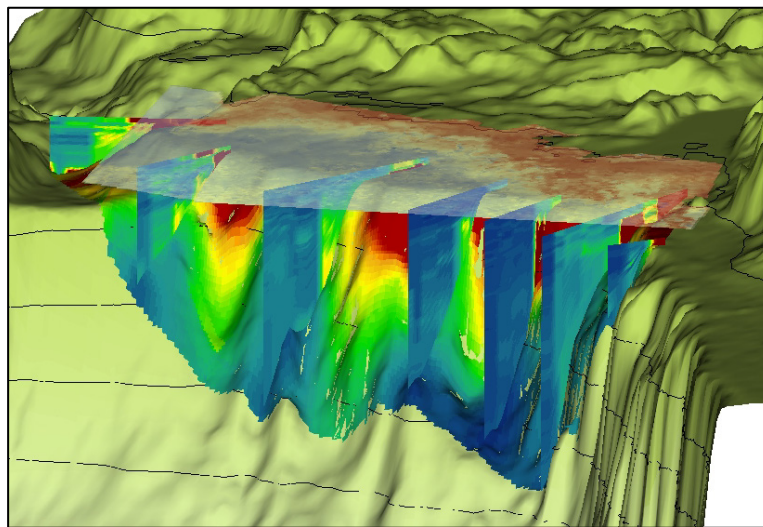


Figure 14. Combinaison des méthodes de visualization : coupes verticales 2D, coupes 2.5D et une image satellite. Le plan 2D horizontal représente le carbone organique particulaire (COP) de surface, généré à partir d'un traitement d'images satellites. Les coupes verticales et la surface 2,5D collée sur le fond bathymétrique sont issues d'un modèle volumétrique du flux de COP dans l'ensemble de la région d'étude (*Ordinary Kriging*). Cette représentation spatiale a été effectuée avec Entervol for ArcGIS (société CTech). Voir également Forest et al. 2013, *matériel complémentaire* pour un modèle interactif de cette image.

CHAPITRE 3 - 3D geospatial modelling and visualization for marine environment: Study of the marine pelagic ecosystem of the south-eastern Beaufort Sea, Canadian Arctic

3.1 Résumé

La modélisation géospatiale de l'écosystème marin pélagique est complexe en raison de sa nature dynamique et volumétrique. Par conséquent, l'analyse spatiale océanographique habituelle de cet environnement est en 2D, limité aux coupes statiques horizontales ou verticales pour représenter différents phénomènes. Dans cet article, nous explorons la contribution du développement récent en SIG 3D et en outils de visualisation scientifique pour la représentation et l'analyse de données océanographiques. Les avantages d'une solution 3D sont illustrés avec une représentation géospatiale 3D de la distribution des masses d'eau dans la partie sud-est de la mer de Beaufort (ouest de l'Arctique canadien).

3.2 Abstract

Geospatial modelling of the marine pelagic ecosystem is challenging due to its dynamic and volumetric nature. Consequently, conventional oceanographic spatial analysis of this environment is in a 2D environment, limited to static cutting planes in horizontal and vertical sections to present various phenomena. In this paper, we explore the contribution of recent 3D development in GIS and in scientific visualization tools for representation and analyses of oceanographic data sets. The advantages of a 3D solution are illustrated with a 3D geospatial voxel representation of water masses distribution in the southeastern Beaufort Sea (west of the Canadian Arctic).

3.3 Introduction

Oceans cover 71% of the Earth's surface and with an average depth of approximately 4000 m, the volume of the marine pelagic ecosystem (water column) represents 99% of the biosphere (Angel, 1993). From a geospatial modelling perspective, this environment imposes different challenges compared to the terrestrial system. Whereas the latter is most often represented as an empty space filled with object (trees, houses etc...), the pelagic ecosystem is a continuous abiotic and biotic spatial geographical phenomenon in a full three-dimensional (3D) environment. Study of the marine ecosystem, as well as management and conservation of marine resources, can be enhanced with adequate geospatial 3D modelling.

Traditional Geographic Information Systems (GIS), that are leading tools for the study and observation of spatial data, are not suitable to model geoscientific datasets since they have been principally designed for static and two-dimensional (2D) objects in terrestrial applications (Carette et al., 2008; Ledoux and Gold, 2008; Wright and Goodchild, 1997). Nevertheless, recent 3D GIS development justifies an update of their potential for marine pelagic geospatial modelling. Although lacking much of the flexibility in data management and ease of use of GIS, various specific scientific visualization tools have also been developed for modelling of the geologic subsoil, mainly motivated by oil and gas industry. These geomodelling tools are interesting from an oceanographer's perspective in that they treat, at least partially, the same type of continuous field as we find in the pelagic ecosystem.

In this paper, we first analyse and compare the capacity of different GIS tools for representation, visualization and analysis of a 3D dynamic marine environment. Specifically, we conduct a qualitative comparison between capacities of commercial GIS, commercial marine GIS, and academic prototype GIS as well as geomodelling tools (section 3.4). We demonstrate how recent development in 3D spatial modelling tools can improve representation, visualization and analyses of oceanographic phenomena and highlight some improvements that should be carried out to these tools in order to achieve an optimal marine spatial modelling tool. Finally, we propose an integration of the benefits from 3D geomodelling tools with advantages of GIS to improve 3D spatial modelling of oceanographic data sets. The third part of the paper is devoted to a case study proposing a 3D solution to visualization of water masses distribution in the southeastern Beaufort Sea (west of the Canadian Arctic). The data for this case study were obtained from the Malina oceanographic campaign conducted over the Mackenzie shelf between the 31th of July and 26th of August in 2009 (Fig 15). More detailed information about the Malina campaign can be found elsewhere (e.g. Matsuoka et al. 2012).

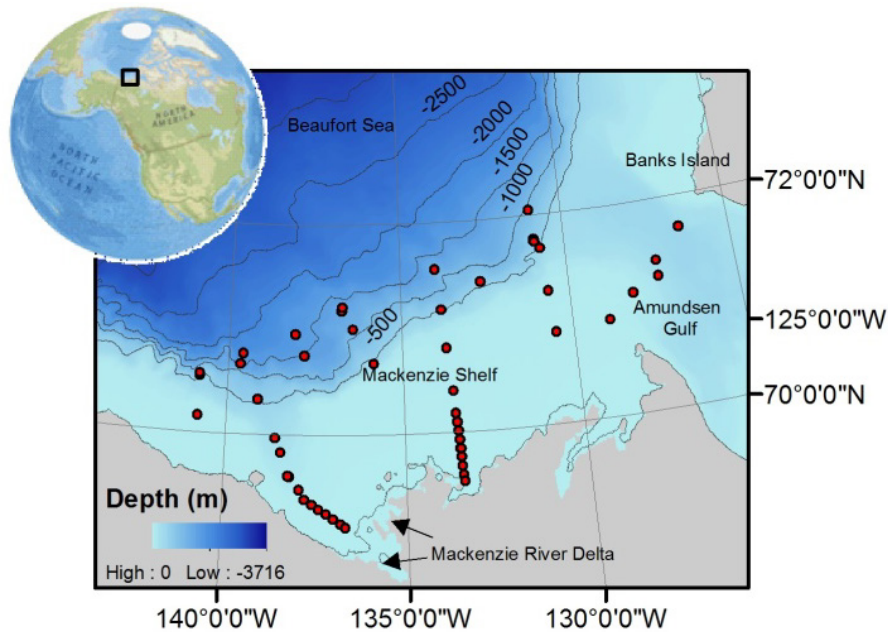


Figure 15. Location of sampling stations for the Malina cruise in the Beaufort Sea, Canadian Arctic.

3.4 3D spatial representation of the pelagic ecosystem

Pelagic marine features are characterized by their fuzzy boundaries, dynamic, and full 3D structure (Gold and Condal, 1995; Shyue and T sai, 1996). These characteristics are restrictive for data acquisition, as well as for geospatial modelling and representation. A further problem with oceanographic data sets is the frequent anisotropic distribution of data, due primarily to logistics and costs associated with expensive sampling at sea. Development of tools available for 3D geospatial modelling of the marine pelagic ecosystem is consequently challenging. This might explain why conventional analysis of oceanographic phenomena is traditionally in a 2D environment, limited to static cutting planes in horizontal and vertical sections either contoured or colour-coded to present various parameters (Head et al., 1997).

Oceanic physical parameters, such as temperature and salinity, define distinct water masses with more or less fuzzy boundaries. Vertical and horizontal distribution of these water masses influences the oceanic carbon cycles, which in turn play an important role in regulating global climate. One partial objective of the Malina oceanographic campaign was to gain better understanding of these interactions. Visualization and spatial analyses in a 3D geospatial model of these phenomena can then be of great value. Such a model could also be of good use in a resource management or conservation perspective.

A summary of some common and specialised tools reviewed in this work are listed in Table 2. These have been evaluated according to criteria for their suitability for 3D geospatial modelling of the pelagic environment.

This review indicates that recent efforts of 3D development in the GIS field have mostly focused on object centred conceptual design using vector structures (for example: ArcGIS version 10, Fledermaus). Indeed, several research teams have recognized the lack of adequate 3D marine GIS (Arsenault et al., 2004; Mesick et al., 2009). We consider that an adequate representation of marine pelagic continuous phenomena needs fully developed volumetric field representations. That kind of representation is more developed in geomodelling tools. The general solution is the use of 3D raster-based models, commonly referred to as voxel (VOlume piXEL) structures. Although in use since the beginning of 1990, they are still mostly absent in commercial GIS (the GIS open-source GRASS might however be an exception to this generality, offering limited volume rendering). It is also worth noticing that the more dynamic data structures -Voronoi tessellations, whose advantages have been stressed in various academic works (Hashemi Beni et al., 2011; Ledoux and Gold, 2008), are absent so far in commercial geospatial modelling tools. A draw-back with the geomodelling tool reviewed in this paper is its limited ability for image treatment and analyses, imperative in oceanographic research considering the common use of remote sensing. Another essential function for an optimal marine geospatial modelling tool is the visualization of 2D static cuts in vertical direction. Even though such a function might seem trivial and does not require a true 3D interpolation method, this operation is not possible at present with ArcGIS and very limited with EnterVol, commercial extension to ArcGIS that permits volumetric representation (Tableau 2). Finally, none of the tools evaluated permit to take into consideration the dynamic nature and fuzzy boundaries of pelagic phenomena or to assess a general predictive capability of spatial 3D models, such as cross-validation (e.g. (Foglia et al. 2007)).

Tableau 2. Review of five geospatial modelling tools from commercial and academic GIS as well as from geomodelling.

Criteria	ArcGIS 10	Fledermaus	HabitatSpace	CTech EnterVol	Paradigm Gocad
	<i>Commercial GIS</i>	<i>Commercial marine GIS</i>	<i>Academic prototype GIS</i>	<i>Extension to ArcGIS</i>	<i>Geomodelling tools</i>
3D Interpolation	-	Limited	(Kriging)	Kriging (limited), IDW	Kriging, IDW, other.
3D raster representation	-	Limited	✓	✓	✓
3D vector representation	✓	✓	✓	✓	✓
Visualization cuts	-	✓	?	Limited	✓
Visualization iso-surfaces	-	✓	✓	✓	✓
Visualization volumes	-	-	Limited	✓	✓
3D statistical analyses	-	-	?	✓	✓
3D spatial analyses	-	-	Limited	✓	✓

3.5 3D geospatial solution to representation of water masses distribution: Malina case study

In the south-eastern Beaufort Sea, several types of water masses can be identified, such as the nutrient rich pacific Upper Halocline Water (UHW). The fractional presence of UHW was obtained for each of the 243 sampling points (x, y, z) accordingly to a method described by Lansard et al. (2012). A geospatial voxel model of this water mass was constructed with Paradigm GOCAD, a scientific visualization tool developed for 3D geological spatial modelling. This spatial model was built by a grid of 150 x 75 x 100 voxels in x, y, z direction and compressed vertically between water surface and bathymetric surface. UHW values were attributed to each voxel in the model through a 3D interpolation of sampling points with ordinary kriging. The kriging variogram's dependent predictive error is presented in Figure 16. In general, the lower the error for a specific location, the better is the prediction of the spatial model.

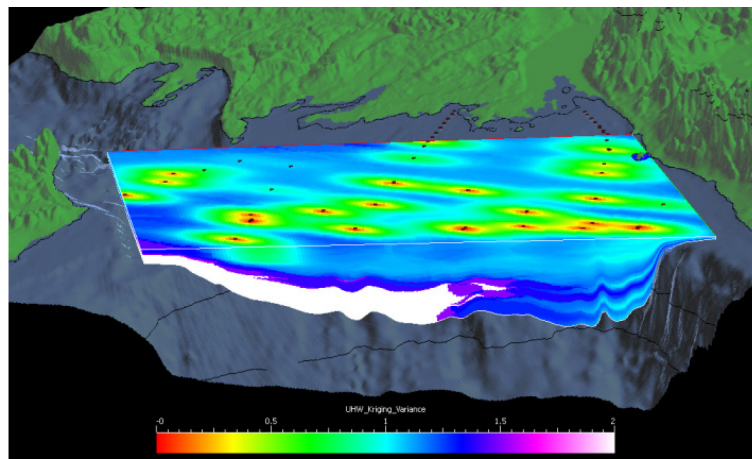


Figure 16. Kriging variance of spatial 3D model for pacific Upper Halocline Water in the Beaufort Sea. Black dots indicate sampling points.

The final spatial model permits us to visualize iso-surfaces and volumes as well as cuts in any plane of the water mass (Figure 17) and enables spatial 3D analyses, such as volume calculation and intersection. This case study shows that recent advances in volumetric representation developed primarily for geomodelling tools can be used to extend usual interpretation of static marine pelagic phenomena from 2D static planes to a 3D environment.

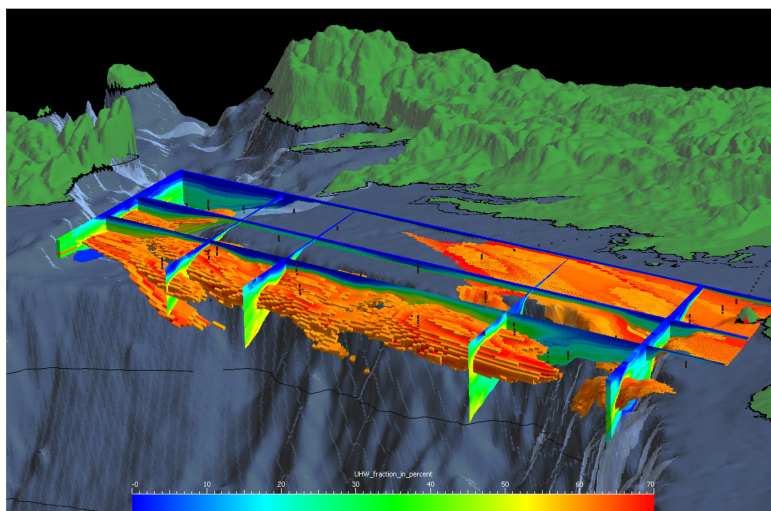


Figure 17. Geospatial representation of Upper Halocline Water on the Mackenzie shelf. Model constructed with Paradigm Gocad). Scale indicate fraction (%) of water constituted of UHW. Voxels containing more than 60 % of UHW are coloured-filled. Black dots indicate sampling points. Vertical exaggeration in figure is 75 times that of reality.

3.6 Discussion and conclusions

This research has explored the potential of 3D geospatial modelling tools for the study of marine pelagic ecosystems. A review of common GIS indicates that these softwares mostly lack the necessary functions for volumetric representation of gradual phenomena, primordial for geospatial study of marine ecosystems. However, performance of geomodelling tools for representation of this environment is promising, which is illustrated in this paper by a snap-shot of a 3D solution to visualization of a water mass in the south-eastern Beaufort Sea, constructed with Paradigm Gocad (Figure 17). Integration of volumetric representation in a GIS environment is an important advance towards an optimal marine GIS. This tool must also include representation and analyses functions of 2D static vertical cuts. On the contrary, geomodelling tools could be adapted to the marine environment by improving basic functions for oceanographic research such as image treatment and analyses. However, all spatial modelling tools conceived for the pelagic environment would also benefit from including spatial data structure that takes into consideration the dynamic nature and fuzzy boundaries of the pelagic environment. Our future research will explore the contribution of 3D geospatial modelling tools for the combination of satellite and ground data in the identification of biological hot-spots and organic carbon fluxes in the Beaufort Sea.

3.7 Acknowledgements

The authors are grateful for the partial financial contribution to this project from Natural Sciences and Engineering Research Council of Canada.

CHAPITRE 4 - Assessment of 3D spatial interpolation methods for study of the marine pelagic environment

4.1 Résumé

Compte tenu de la nature volumétrique de l'océan, la modélisation et l'interpolation spatiale 3D pourraient être la clé d'une meilleure compréhension des phénomènes abiotiques et biotiques continus de l'écosystème marin, bien que ces techniques soient rarement utilisées et leurs performances réelles mal étudiées. Nous évaluons ici la performance de l'interpolation 3D pour cinq variables pélagiques marines issues d'une campagne océanographique typique, réalisée dans la partie sud-est de la mer de Beaufort (Arctique canadien) en 2009. Notre objectif principal était d'évaluer et de comparer la performance d'une méthode d'interpolation déterministe (distance inverse : IDW) et une méthode géostatistique (ordinary kriging : OK) avec une variation du paramétrage (voisinage, poids) pour des variables d'une complexité croissante en termes d'anisotropie de données et de configuration d'échantillonnage. La performance des stratégies d'interpolation est évaluée par une validation croisée et par une comparaison qualitative visuelle. Cette étude montre que l'OK était la méthode optimale. Cependant, lorsque la complexité des variables pélagiques augmentait en termes d'autocorrélation spatiale et de variance de données (p.ex., variables avec une plus grande complexité au niveau de la fonction biologique), la différence de performance entre OK et IDW diminuait. Nous recommandons que les progrès récents en matière de modélisation spatiale 3D, développée principalement pour des outils de modélisation géologique, soient exploités pour étendre l'interprétation habituelle des phénomènes pélagiques marins 2D à un environnement 3D.

4.2 Abstract

Given the volumetric nature of the ocean, 3D spatial modeling and interpolation could be key to a better understanding of continuous abiotic and biotic phenomena that compose the marine ecosystem, although such techniques are rarely used and their actual performance are poorly studied. Here, we evaluate the performance of 3D spatial interpolation for five pelagic variables derived from a typical oceanographic campaign, conducted in the south-eastern Beaufort Sea (Canadian Arctic) in 2009. Our main objective is to evaluate and compare the performance of a deterministic interpolation method (inverse distance: IDW) and a geostatistical method (ordinary kriging: OK) with a variation of method input parameters (searching neighborhood, weighting) for variables with increasing complexity in terms of data anisotropy and sampling configuration. Performance of different interpolation strategies is evaluated by cross-validation and a qualitative visual comparison. Our results illustrate that OK was the optimal method. However, when complexity of pelagic variables increased in terms of spatial autocorrelation and data variation (i.e. biologically-related variables), error difference between OK and IDW was reduced. We recommend that recent advances in spatial 3D modeling tools developed primarily for geological modeling should be exploited to extend the usual interpretation of marine pelagic phenomena from a 2D to a 3D environment.

4.3 Introduction

Oceans cover 71% of the Earth's surface and with an average depth of approximately 4000 m, the volume of the marine pelagic ecosystem (water column) represents 99% of the biosphere (Angel 1993). This ecosystem, the largest on earth, is a spatial and geographical continuum of abiotic and biotic phenomena in a full dynamic three-dimensional (3D) environment. The study of the vertical and horizontal distribution of oceanic pelagic variables, such as temperature (T), water masses and chlorophyll concentration (chl) constitutes the very basics for any oceanographic research.

The traditional way of acquiring marine pelagic data in coastal zones is an *in situ* sampling strategy along oceanographic transects perpendicular to continental shelves. Due primarily to logistics and costs associated with expensive sampling at sea, sampling stations are sparsely distributed both spatially and temporally. Estimation of values at un-sampled locations (spatial interpolation) is accordingly important for the representation and understanding of different spatial phenomena distribution. Conventionally, spatial representation of geographic phenomena has been done in a 2D environment, limited to static cutting planes in horizontal and vertical sections either contoured or colour-coded (Carette 2008; Carette, Mostafavi, et Devillers 2008; Head et al. 1997). However, limiting analyses to a 2D environment inevitably induces a loss of spatial information as pelagic variables actually fill a 3D space. 3D spatial modeling permits the flexibility to create vertical and horizontal cutting planes in any azimuth and can hence provide a better insight to spatial variations than 2D representations. The 3D spatial model also gives a more comprehensive picture of phenomena distribution as irregular sample points (that is: not only sample points along straight cruise transects) can be included in analyses. Full 3D representations can be used to delimit phenomena such as upwelling (currents bringing up deeper water towards surface) and other hot-spots for biological production. Besides, recent technological advances in data acquisition (such as satellite remote sensing, remotely controlled underwater vehicles) have multiplied the amount and type of data available in oceanographic research and integration of such data further increases the necessity of extending analyses to a full 3D environment. This shift requires robust and performing spatial interpolation techniques.

The search for an optimal interpolation method for non-discrete data is an ongoing research field and published research is not unequivocal on which method is most accurate (Li et Heap 2011; Li et al. 2011; Zimmerman et al. 1999). Many studies in diverse environmental sciences have assessed and compared different interpolation techniques in a 2D space (Daly 2006; Li et Heap 2008, 2011; Li et al. 2011; Li et al. 2010; Mueller et al. 2004; Zimmerman et al. 1999). However, few comparative studies have been performed in a full 3D space and these are mostly related to geological modeling related to mining activities (Falivene, Cabrera, et Saez 2007; Falivene et al. 2010; Jones et al. 2009). While a few comparative studies exist for

marine features in a 2D space (Li et al. 2010), none has been identified that assesses performance of interpolation techniques in a pelagic (3D) environment. This research represents an attempt to examine and validate the performance of 3D spatial interpolation techniques for the marine pelagic environment. It is thus an important step towards the development of a full marine 3D GIS.

Pelagic marine features are characterized by their fuzzy boundaries, dynamic, and full 3D structure (Gold et Condal 1995; Shyue et Tsai 1996). These characteristics make data acquisition, as well as spatial interpolation and modeling, difficult. Already in a 2D modeling context, spatial interpolation is one of the most challenging issues in geosciences (Babak et Deutsch 2009). In the marine domain, the complex layering of pelagic variables along the vertical axis when compared with the horizontal plane also adds to the challenge of 3D interpolation. Geographic Information Systems (GIS) that are leading tools for the study and analysis of spatial data, have not so far been adapted enough to the study of the marine pelagic environment. These have principally been designed for modeling of static and two-dimensional (2D) objects in terrestrial applications (Carette et al., 2008; Ledoux and Gold, 2008). Furthermore, recent efforts in 3D GIS development have focused on vector structures that are not well adapted for modeling of the continuous phenomena of the pelagic environment: an adequate representation of marine pelagic phenomena needs fully-developed volumetric field representations. Such data structures are largely absent from commercial GIS, but are used in geomodelling tools that treat, at least partially, the same type of continuous field as we find in the pelagic ecosystem (Ledoux et Gold 2006; Ledoux et Gold 2008; Sahlin et al. 2012). The general solution is the use of 3D raster-based models, commonly referred to as voxel (VOlume piXEL) structures. A prerequisite for constructing such structures is spatial interpolation. The objective of spatial interpolation is to calculate estimates of variables at unsampled locations by some sort of weighted average of nearby data (Babak et Deutsch 2009; Weber et Englund 1994).

Here, we compare the performance and analyze the robustness of 3D interpolation strategies in raster-based models (voxel models) for the pelagic ecosystem of the Mackenzie Shelf, located in the south-eastern Beaufort Sea (west of the Canadian Arctic). We first briefly introduce the classification of interpolation methods assessed with a special focus on 3D interpolation and 3D variogram modeling. In the following sections, we assess the performance of Ordinary Kriging (OK) and two variants of Inverse Distance Weighting (IDW) with the objective to identify a *robust* and *optimal* 3D interpolation strategy (terms adopted from Falivene et al. 2007). An *optimal* interpolation method within this paper refers to a strategy with minimized estimation errors whereas *robust* refers to a method with estimation errors that are stable with respect to variation of method input parameters (search conditions, weight parameters). *Interpolation strategy* is used throughout this paper to indicate an interpolation method with a specific combination of input parameters. Estimation errors are

assessed using cross validation techniques (CV) and by a qualitative examination of derived geospatial models. Each interpolation strategy is assessed for five pelagic variables with increasing complexity regarding spatial autocorrelation, sampling configuration and ecological function (e.g. relation with nutrient element cycles). These variables are: 1) temperature (T); 2) chlorophyll *a* concentration (referred to as chl *a*); 3) beam attenuation coefficient (c_p); 4) distribution of upper halocline water mass distribution (UHW) and 5) vertical flux of particulate organic carbon (POC-flux). In the last part of this paper, we discuss results and demonstrate how an optimal parameter setting improves spatial 3D interpolation and representation of oceanographic phenomena. Final optimal spatial 3D representations are presented for Beaufort Sea and ecological significance is discussed.

4.4 Background

Spatial interpolation techniques differ from other modeling approaches in that they incorporate information on the geographic position of the sample data points (Babak et Deutsch 2009). In doing so, they offer means of characterizing a variety of factors over different spatial scales (Schloeder, Zimmerman, et Jacobs 2001).

Common spatial interpolation techniques can be largely classified into two groups. If no assessment of errors with the predicted values is performed, the method is considered as being non-geostatistical or deterministic, as weight is assigned according to a deterministic criterion (Li et Heap 2008; Li et al. 2011). When a statistical criterion (spatial covariance) is used for assigning weights, the interpolation method is considered as being geostatistical (stochastic). The field is then considered as a random process and the optimality of the averaging method is determined by minimizing the estimation variance (Babak et Deutsch 2009).

IDW with its variations, is one of the most popular deterministic methods, whereas OK is one of the most extensively used geostatistical methods (Babak et Deutsch 2009; Li et Heap 2008; Li et al. 2010). Both methods produce interpolations that are weighted averages of the observed data and can thus be described by the equation:

$$\hat{z}(u) = \sum_{i=1}^n w_i(u) z_i \quad (8)$$

where $\hat{z}(u)$ represents the estimated value at the location u , $w_i(u)$ is the weight at location u for the i^{th} sample, z_i denotes the measured value and n denotes the number of neighboring points used for the summation.

IDW and OK have been extensively compared in 2D spatial environmental assessments (Kravchenko et Bullock 1999; Li et Heap 2008, 2011; MacCormack, Brodeur, et Eyles 2013; Mueller et al. 2004; Schloeder, Zimmerman, et Jacobs 2001; Zimmerman et al. 1999). It is recognized that method performance is data- and variable specific (Cressie et Zimmerman 1992; Li et al. 2011; MacCormack, Brodeur, et Eyles 2013; Zimmerman et al. 1999). Method input parameters (e.g. search neighborhood, method specific parameters)

are also important, but no agreement exists on how these parameters should be set (Bond, Gibbs, et Jones 2007; Li et al. 2011; MacCormack, Brodeur, et Eyles 2013), even though methods such as data transformation, declustering and quantitative kriging neighborhood analysis (Deutsch 2002; Kelkar et Perez 2002; Olea 2007; Vann, Jackson, et Bertoli 2003) assist the users in making informed parameter choices prior to estimation. It has been shown that performance of IDW is sensitive to the number of neighboring points (n) that is used in the estimation (Babak et Deutsch 2009; Kravchenko et Bullock 1999; Zimmerman et al. 1999), whereas kriging methods would show a certain robustness with regard to this parameter (Babak et Deutsch 2008; Weber et Englund 1992). The importance of the exponential parameter (*power*) of IDW has also been stressed (Babak et Deutsch 2009; Falivene et al. 2010; Li et al. 2008; Li et al. 2010; Weber et Englund 1992).

4.4.1 Inverse Distance Weighting

IDW is extensively used because of its ease of implementation: it constitutes a straightforward method with low computational load that is available in most GIS and other scientific visualization software. In IDW, the weights for a location u are inversely proportional to the distance or some power of the distance between the data point being estimated and the observed values i ;

$$w_i(u) = \frac{\frac{1}{d_i^p}}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (2)$$

where d is the distance from the i^{th} sample location to the estimation point. The power p is defined by the user and as p increases, the weight of distant points diminishes. Commonly, an inverse distance squared weighting is employed (p is set to 2). When $p = 0$ and $n = 1$, each estimation point is taking the full value of the nearest sampling point. This specific case of IDW is also called Nearest Neighbor interpolation (NN) and is used in this research as a control of interpolation method performance. The use of NN as a baseline method is justified for several reasons. First, NN does not require any parameter assumptions and potential subjective interactions are hence minimized. Secondly, pelagic variables are characterized by smooth gradual changes (absence of abrupt fluctuations) (Lansard et al. 2012; Minster 1997), which implies that the local value is normally close to the general (local) tendencies. In such a context, NN can be assumed to produce reasonable estimates.

A variant of IDW where the Euclidean xyz -space is replaced by an uvw -grid space was also assessed (from here on, these methods are referred to as IDW_{xyz} and IDW_{uvw} , or jointly IDW-methods). Using an uvw -grid space implied an exaggeration of coordinates in vertical direction, as dimension of grid cells had been designed to capture the approximate data variation of variables; one lag in z -direction (5-25 meters, depending on variable) was equalled with a lag in horizontal layer (1935 - 4900 m). Consequently, this grid space implied a vertical exaggeration (length of grid cell in horizontal direction / length of grid cell in vertical direction) by a

factor varying from 111 (T) to 800 (chl a , c_p). The IDW_{uvw} approach was justified by the usual sample configuration in an oceanographic context (dense vertical and sparse horizontal sampling) combined with the normal stratification of continuous variables in the vertical plane. As IDW interpolation uses search criteria where n sampling points will be weighted according to their distance to the arbitrary point u at which the estimation is carried out, the majority of n closest points to u will inevitably be derived from one or a few vertical sampling stations (CTD-casts). This is undesirable when considering the natural stratification of pelagic variables.

4.4.2 Ordinary Kriging and 3D variogram

Kriging interpolation methods have been developed within mining engineering in the subfield of statistics named geostatistics. The advantage of geostatistical methods resides in their capacity to incorporate knowledge of underlying spatial structure of the phenomenon of interest into unbiased estimates (Reed, Ellsworth, et Minsker 2004). Briefly, both the search range specifications and the weighting function are based on analyses of semi-variograms (from here on simply referred to as *variogram*). Kriging is therefore taking into account equally distance and direction of spatial autocorrelation for the estimation and provides a measure of accuracy of interpolated values (*kriging values*), and consequently the reliability of the geospatial model. However, the subjectivity of variogram modeling is often emphasized and constitutes a subject for an ongoing debate in the geostatistical scientific community.

In OK, the mean of the variable is assumed to be stationary and unknown while the kriging estimator is designed to provide the best linear unbiased estimate. Accordingly, it is considered as a statistical *optimal* interpolator as it minimizes the estimation variance (mean squared prediction error; MSE) at each location to be estimated (Babak et Deutsch 2008; Cressie et Zimmerman 1992; Weber et Englund 1992). For every estimate \hat{z} , a local mean is calculated by using the sample data points found within the neighborhood specified. The OK equation has been described in detail elsewhere (Cressie et Zimmerman 1992; Isaaks et Srivastava 1992) and will not be repeated here. The major difference between 2D and 3D kriging is related to how data anisotropy is treated by the variogram modeling.

The *experimental* variogram at a given interval distance consists of average squared differences of data taken at sites lagged by that distance. Variogram modeling in 1D and 2D environment has been extensively defined in literature and readers are referred to reference papers for precise equations (Goovaerts 1997; Isaaks et Srivastava 1988). In 3D, three angles are used to define the direction vector of the anisotropic experimental variogram: azimuth, dip and plunge (Dahle, Kolbjornsen, et Abrahamsen 2005; Desassis et Renard 2012; Gringarten et Deutsch 2001; Mucha et Wasilewska-Blaszczyk 2012). The azimuth angle represents the clockwise rotation of the vector in the horizontal plane, the dip angle the downward rotation from the horizontal

plane and the plunge angle the rotation of the minor horizontal direction counterclockwise relative to the principal direction. A set of experimental variograms in different azimuth and dip angles are generated until the three principal axes (X, Y, Z) of anisotropy are found. These axes are identified from the combination of experimental variograms that gives the greatest difference in spatial range of autocorrelation (*range*). A valid variogram model (*theoretical* variogram) is selected and fitted to the set of experimental variograms. Three ranges (major and minor in horizontal plane, as well as vertical range) are derived from the set of experimental variograms and used in the calculation of the kriging weights.

When data density is too sparse to permit variogram modeling in different dip angles, a common way to perform 3D variogram analyses is to model vertical (1D variogram in vertical direction) and horizontal variograms (2D variogram) separately (Dahle, Kolbjornsen, et Abrahamsen 2005; Falivene, Cabrera, et Saez 2007; Gringarten et Deutsch 2001). In this research, experimental variograms were modeled simultaneously in four azimuths and three dip angles. This approach permits to preserve the relation between azimuths as the variance contributions identified for one direction will hold for other directions (Gringarten and Deutsch 2001; Dahle et al. 2005). It also permits to estimate dip-angles of highest spatial autocorrelation.

4.4.3 Accounting for the dynamic nature of the marine pelagic environment

The dynamic nature of the marine pelagic environment poses a challenge to all spatial representation or modeling, whether it is for 2D cutting planes or a full 3D spatial modeling. In a 3D environment where water masses are in constant movement, the static assumption of variables throughout the duration of the sampling period will always be more or less flawed. Indeed, for the 38 days of the ArcticNet-Malina oceanographic campaign, a significant part of variables' variation may be attributed to their temporal evolution. However, sampling date is only one amongst other factors (abiotic and biotic) that influence values of pelagic variables: in the case of vertical POC-flux during the ArcticNet-Malina campaign, this factor explained only 4.7 % of the total variation whereas the trend due to biotic and abiotic variables explained 13 % (Forest et al. 2013). The non-temporal trends will at least partly be accounted for in the spatial modeling, as various biotic and abiotic factors are also functions (at least partially) of their spatial location. This is less the case for climatic factors, such as the presence of upwelling-favorable winds that was one of the most important factors for biological-related variables such as chl *a*, *c_p* and vertical POC-flux during the Malina campaign (Forest et al. 2013). Even though temporal variability will account for a part of the observed spatial variability in this study, major spatial structures observed will more likely be related to regional differences. Regarding UHW, temporal variability is slow-occurring, taking place on periods longer than the Malina campaign (Lansard et al. 2012). Variation of sea-surface temperature is expected during the time span of the ArcticNet-Malina campaign, but would be limited to the upper first layer of grid-cells: below surface layer, sea-temperature is more closely related to the slow-changing water masses distribution. Finally, major spatial structures of chl *a* and *c_p* are related to a

subsurface chlorophyll maxima (SCM) that is ubiquitous during the ice-free season in Arctic waters (Martin et al. 2010).

Nevertheless, all static representation of oceanographic phenomena will have limitations due to this somewhat flawed assumption of a static environment. Hence, discussion of modeling results should be considered with respect to what part of the data variation that can be attributed to a spatial variation of variables and what fraction that can be attributed to their temporal variation. However, as long as these issues are discussed, the resultant spatial model can be considered as some kind of averaged snap-shot of the reality during the sampling period.

4.5 Methods

4.5.1 Study area and data preparation

4.5.1.1 Sampling design

For the purpose of this research, our data (variables T, chl *a*, *c_p*; UHW and vertical POC-flux) were obtained from the ArcticNet-Malina oceanographic campaign that was conducted on the Mackenzie Shelf in the southern Beaufort Sea from 18th July to 30th August, 2009 aboard the Canadian icebreaker CCGS Amundsen. One of the general objectives of this campaign was to gain an understanding of how the spatio-temporal variability of abiotic factors (e.g. light, nutrients) and biotic factors (e.g. primary production, microbial diversity) influences the dynamic of carbon and nitrogen cycles in the pelagic ecosystem of the Arctic Ocean. The presence of dispersed multi-year ice floes from the decaying Arctic pack as well as the travel distance over an extensive study zone (530 km wide and 270 km wide) largely conditioned sampling design. During the first ten days (ArcticNet campaign), irregularly distributed sampling was concentrated in an area situated in the center of the Mackenzie continental shelfbreak, whereas August sampling (Malina campaign) was performed principally along transects perpendicular to the coastline (Figure 18).

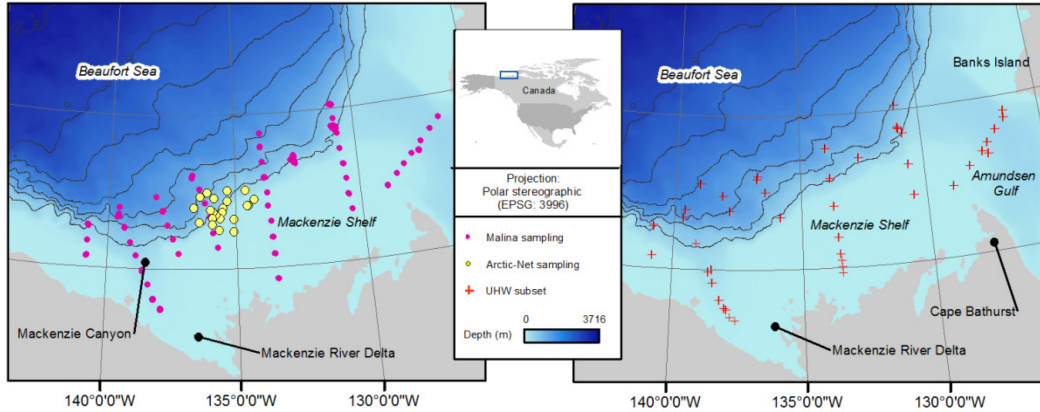


Figure 18. Location of sampling stations used for the spatial modeling of pelagic variables as visited during the ArcticNet-Malina oceanographic campaign, Beaufort Sea (Canadian Arctic) in 2009.

Temperature, fluorescence and c_p were sampled once or several times (one or several casts) at each of the 74 sampling stations using a caged rosette profiler equipped with a conductivity-temperature-depth system (CTD) connected to a fluorometer and a transmissometer. Fluorescence data were calibrated against *in situ* chl *a* concentrations to allow accurate estimations of chl *a* (see Forest et al. 2013 for details). An Underwater Vision Profiler 5 (UVP5) was used to record the particle size-distributions of both non-living particles and zooplankton. By linking these particle distributions to mass and POC-fluxes estimated by sediment traps, POC-fluxes could be estimated on an approximately 25-m vertical interval. Full details of the technical specifications and operations of the UVP5 processing are described elsewhere (Forest et al. 2012; Picheral et al. 2010), while the methodology for transforming these data to POC-fluxes is described in Forest et al. (2013). The fraction of water corresponding to UHW was derived from an optimum multiparameter analysis of salinity, total alkalinity and isotopic composition ($\delta^{18}\text{O}$) of seawater. Further details about this methodology are found in Lansard et al. (2012). Data for UHW were available from a limited subset of the Malina data set (290 xyz sampling points), but could be enhanced by three additional stations sampled from 29th to 30th August during a Geotraces oceanographic campaign (Francois et al. 2009).

4.5.1.2 Data preparation

General characteristics of the data sets are presented in Tableau 3. Two different data sets were used for modeling of temperature. The first one consisted of a subset obtained from the ArcticNet campaign conducted from 16th to 25th July, while the second one consisted of the full ArcticNet-Malina data set. Spatial modeling of chl *a* and c_p was carried out from a data set averaged to a 5 m vertical spacing. As these variables are usually only relevant for studying the light penetrating zone (*euphotic* zone), data were limited to the upper 200 m of the water column. Vertical sample density of POC-flux and UHW was not up-scaled as these variables were

available on a 25 m vertical interval scale (POC-flux) or less (UHW). Spatial modeling of UHW was constrained to a maximum grid depth of 500 m (values indicated a total absence of UHW on depth greater than 450 m). Outliers suspected to contain errors (Tableau 3) were deleted from data sets prior to interpolation. In the case of POC-flux, the 17 superior values (above 1400 mg C m⁻² d⁻²) were excluded. These were all exceeding the data interval for which the power-law algorithm (Forest et al. 2013) linking measured particle size-distribution to sediment trap measurements had been calibrated. Regarding chl *a*, values above 10 mg m⁻³ were excluded as they originated from a process known as "photochemical quenching" that increases artificially and momentarily the fluorescence emitted by phytoplankton cells (thus resulting in unrealistically high chl *a* records). Arithmetic and geometric means (and standard deviation) are presented for variables as outliers were frequent (use of geometric means diminish the influence of outliers).

Tableau 3. Basic statistics for variables and data sets used in research.

	T _{subset} (°C)	T (°C)	UHW (%)	chl <i>a</i> (mg m ⁻³)	c _p (m ⁻¹)	POC flux (mg C m ⁻² d ⁻²)
Data set	ArcticNet (station 1-25)	ArcticNet-Malina	Malina and GEOTRACES	ArcticNet-Malina	ArcticNet- Malina	ArcticNet- Malina
Depth (m)	1-1700	1-1700	0-500	0-200	0-200	1-1700
Number of stations	24	74	42	74	74	74
Samples	9369	2446	290	6047	6047	2239
Vertical sample distance (m)	1	25	70*	5	5	25
Outliers	None	None	None	> 10	None	> 1136
Value range	(-) 1.44-1.99	(-)1.44 - 8.38	0 - 100	0.0376 - 9.07	0.456 - 3.30	1.10 - 1.13 x 10 ³
Median	0.15	(-) 0.055	0.021	0.062	0.48	14.3
Arithmetic mean	(-) 0.17	(-) 0.17	20	0.34	0.51	39.2
Geometric mean	(-) 0.30	(-) 0.34	18	0.0921	0.457	16.7
Arithmetic St. Dev.	0.75	1.12	30.8	0.4150	0.135	99.3
Geometric St. Dev.	1.35	1.43	6.64	2.24	1.16	3.08

* Average vertical spacing of samples.

4.5.2 Computational details

4.5.2.1 Grid design

All variogram modeling and interpolations were performed with the commercial geospatial modeling tool Paradigm GOCAD® (version 2009.2). Size and number of grid cells were adapted to data density in vertical and horizontal directions as to approximate minimal projected distances between sampling points on each axis. Grids were constrained vertically by sea surface and bottom topography obtained from the International Bathymetric Chart of the Arctic Ocean (IBCAO) at a resolution of 2 km (Jakobsson et al. 2008). Grid cells were aligned horizontally with Mackenzie shelf-break and parallel to sea surface with a constant vertical height. Excess voxels were truncated from bottom (Figure 19).

An alternative non-linear grid design was assessed for UHW with where cells that were proportionally compressed between surface and sea-bottom (no truncation of voxels). This grid-design was justified by the fact that spatial distribution of water masses in some oceanographic regions is parallel to bottom topography (aligned along density gradients). However, CV results indicated high errors when using this grid compared to the parallel grid and accordingly, results from the assessment of this alternative grid design will not be any further discussed in this article.

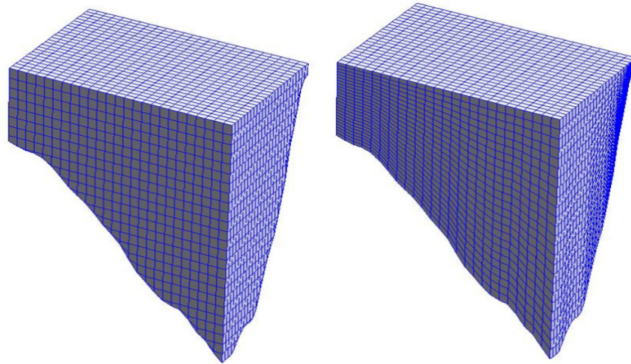


Figure 19. Grid designs assessed. Principal grid design used in study (left) and alternative grid design (right).

4.5.2.2 Variogram settings

The experimental variograms were always constructed with the entire data sets. Standardized variables (z-score transformation) were used for constructing experimental variograms: $Y = (Z - m) / \sigma$ (where m and σ are the mean and the standard deviation of the observed variable Z). This approach introduces a major advantage for variogram analyses as the sill is fixed to 1 in the theoretical model (the variance of the transformed variable = 1), which also increases comparison possibilities of spatial structures between variables. Ranges of autocorrelation and major structures of variogram from transformed variables were verified and corresponded well with previous spatial mapping on the MacKenzie Shelf of the pelagic variables used in this study (Lansard et al. 2012; Martin et al. 2010; Forest et al. 2013). Transformed data were used only in variogram modeling. Estimation process was continued using the variogram results and non-transformed data (in this case, there is no need for back-transformation). 3D variograms were constructed for all variables in four horizontal intervals of azimuths (central angle 0, 45, 90, 135 degrees) and three intervals of dip angles (central angle 0, 45 and 85 degrees, where 0 = horizontal plane) for anisotropy analysis of the behavior of the variables. Plunge angle was always kept zero. This strategy yielded a set of 12 interrelated experimental variograms for each variable (illustrated for c_p in Figure 19). The following basic modeling conditions were used: (1) the same mathematical function was used in all twelve directions, (2) the nugget was estimated from variogram in the azimuth and dip angle holding the lowest nugget value (in our case, always vertical direction), as proposed by Gringarten and Deutsch (2001), (3) The sill was hold constant to 1 (standardized variable) and (4) the nugget had the same

value in all directions. A *spherical* theoretical model was fitted to all experimental variograms, as this model visually was the optimal choice. No assessment of different theoretical models was carried out as an evaluation of robustness of variogram was carried out by varying derived ranges (see next section).

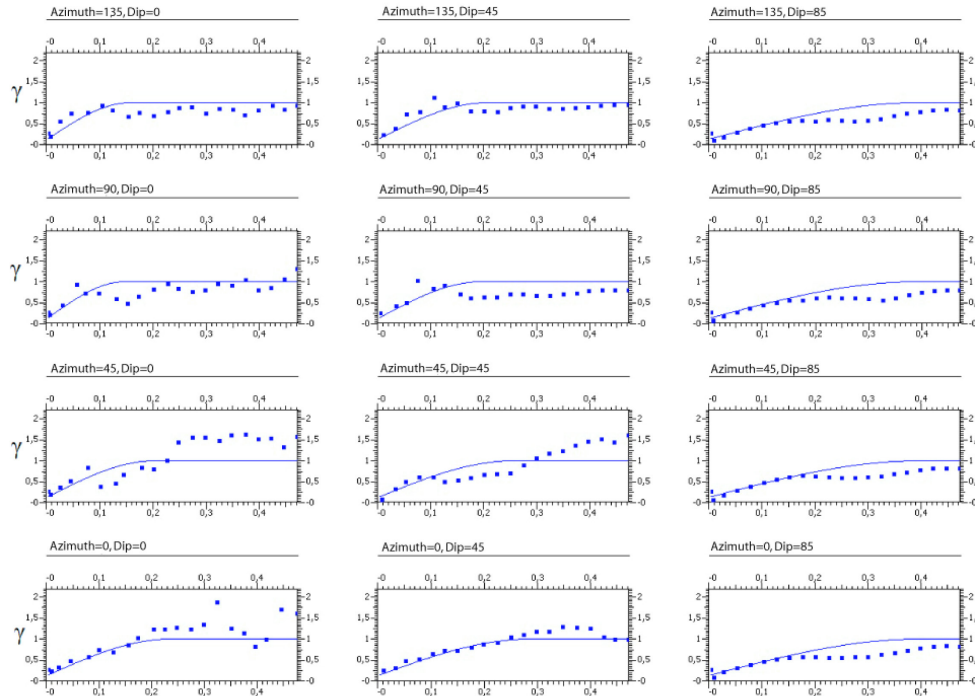


Figure 20. Anisotropic 3D variogram for c_p with fitted spherical model. Y-axes correspond to semi-variogram variance (γ), while X-axes correspond to distance in grid units. The twelve interrelated variograms are used to identify directions (azimuth, dip) of minimum and maximum spatial continuity. Theoretical ranges of spatial autocorrelation is obtained from the X-axes where theoretical variogram model (solid line) reaches a sill value of 1 ($\gamma = 1$).

4.5.2.3 Parameter setting for interpolation strategies

In order to verify the robustness of interpolation strategies and identify an optimal parameter setting, different combinations of input parameters were used with regards to search conditions and weight of neighboring data (Tableau 4). For IDW methods, powers (p) of 1, 2 and 5 were assessed. The quite high value of $p = 5$ was justified by the spatial distribution of variables, sampling configuration and preliminary tests on different powers indicating acceptable estimations for this value. The sampling configuration (dense vertical and sparse horizontal sampling) implied that a majority of sampling points (n) weighted would expectedly be located in one or only a few vertical sampling stations (CTD-casts). Weighting of many sampling points from the same station could be undesirable when considering the normal stratification of pelagic continuous variables in the vertical plane. As a higher p increases local (thus decreases influence from distant vertical locations), an assessment of a higher p was relevant. The robustness of OK interpolation method was assessed by varying horizontal

ranges (derived from 3D variogram analyses) with an underestimation or exaggeration by a factor of 1.5 (vertical range was always kept stable). This factor seemed reasonable from examination of raw data, variogram results, geographical size of study zone and previous oceanographic studies on MacKenzie Shelf (Forest et al. 2013; Lansard et al. 2012; Martin et al. 2010). The optimal number of neighboring data (n) was assessed by varying this parameter in the range of 8 to 96. A somewhat lower number of n was used for assessment of UHW due to the limited number of sampling points for this variable. No assessment of nugget was performed, as nugget-to-sill ratios were low for all variables (Tableau 4) and assumed to be relatively precise due to the dense sampling in vertical direction.

Tableau 4. Variation of input parameters for the assessment of robustness of interpolation strategies. n refers to the number of sampling points used for the weighted estimation, *horizontal ranges* refer to distances of spatial autocorrelation in the horizontal plane derived from variogram modeling and p corresponds to the IDW-power.

Method	Parameter	Variable		
		T _{subset}	T, chl a , c_p , POC flux	UHW
OK	n	16, 32, 48, 96	16, 32, 48, 96	8, 16, 32
	Horizontal ranges	Underestimation or exaggeration by a factor of 1.5*		
IDW _{xyz} , IDW _{uvw}	n	16, 48, 96	8, 16, 48, 96	8, 16, 32
	p (power)	1, 2, 5		

* Underestimation of POC flux by a factor of 1.5 yielded kriging variances that indicated that OK was not appropriate for this variable (see Result section). It was hence not considered in further comparisons of results.

Negative weights of OK resulted in some estimations (<1% of values) below physical limits (T, UHW) or detection values (chl a , c_p , POC-flux). Such values were corrected to minimum observed value of data range. When no data points were found within a constrained search neighborhood (OK), the unknown value cannot be estimated and a NDV (no data value) is returned for the voxel. Voxels that were assigned a NDV with one of the OK strategies were eliminated before CV assessment for all interpolation strategies of the same variable.

4.5.3 Method performance comparison criteria

IDW and OK methods performance was assessed with a 3D version of CV. In a 2D environment, the common practice is to remove one sample point at the time from the data set. The interpolation at the sample location is then carried out without the sample and the difference between estimated (\hat{z}) and observed value (z) is calculated. This procedure is repeated for the entire data set. In 3D geospatial modeling of well data, a procedure where an entire vertical column of sampling points is extracted at the time has been suggested (Deutsch 2002; Falivene, Cabrera, et Saez 2007; Kelkar et Perez 2002). Once again, prediction of the extracted points is performed with remnant data set, but this time for the entire column. As described previously, the ArcticNet-Malina campaign produced data sets that share the non-homogenous sample

distribution common for geological well data (scarce sampling horizontally and dense vertically). However, a repeated sampling at many oceanographic stations, important especially during Malina sampling period (31th July to 24th August), created an additional clustering in the horizontal plane: while one vertical sampling (*cast*) per station was generally performed during ArcticNet campaign (18th to 27th July) an average number of 3.2 casts per station was performed during the Malina campaign (*sampling station* will be used from here on to represent an individual cast (T_{subset} , UHW) or a group of casts (T , c_p , chl a and POC-flux)). As the CV process is known to create an underestimation of errors when stations are located in pairs or clusters (Daly 2006), CV assessment was performed by extracting all casts belonging to the same sampling station from the data set for variables T , c_p , chl a and POC-flux. However, this was not necessary, for T_{subset} and UHW data sets (Figure 18) as clustering was considered negligible for these variables.

In summary, the CV process used in this research consisted of:

- 1) removing temporarily all sampling points associated with one sampling station from data set;
- 2) estimating at sampling station extracted under step 1 with remaining data points;
- 3) calculating CV errors for all data points extracted (that is, comparing measured values to estimated values);
- 4) calculating an average CV error for data points extracted;
- 5) repeating steps 1-4 for each sampling station;
- 6) calculating an average CV error for each interpolation strategy.

The CV error definition was assessed using mean absolute error (MAE) and mean squared error (MSE):

$$\text{MAE} = \frac{1}{o} \sum_{j=1}^o \left[\frac{1}{n} \sum_{i=1}^n |\hat{z}_{ij} - z_{ij}| \right] \quad (11)$$

$$\text{MSE} = \frac{1}{o} \sum_{j=1}^o \left[\frac{1}{n} \sum_{i=1}^n (\hat{z}_{ij} - z_{ij})^2 \right] \quad (12)$$

where z_j and \hat{z}_{ij} are the measured value and the estimation value at the location i for the sampling station j , n represents the number of sampling points in the j^{th} -station temporarily removed for the CV and o represents the number of sampling stations for each variable.

These two CV error definitions are commonly used (Willmott et Matsuura 2005) and arguably the most accurate (Li et al. 2008). In general, MAE indicates the extent to which the measure can be in error, whereas MSE measures statistical precision due to both variance and bias. MSE measures the central tendency and

detects extreme errors, being sensitive to the presence of outliers in data. CV results for IDW_{xyz}, IDW_{uvw} and OK were compared to those of Nearest Neighbor (NN) interpolation, used as a quality control.

A general drawback with the CV-error estimation is that no error information is given for geographical locations where there are no sampling stations (Daly 2006). Geostatistical interpolation methods will provide a precision of the accuracy of an estimated value (kriging variance), but no actual error of the value measured. Hence, with no other validation in this context, the qualitative inspection of spatial models is an important part of the validation process (Daly 2006). This evaluation is subjected to the researcher's knowledge of the physical environment and was performed for each interpolation strategy with optimized parameters.

4.6 Results

Ordinary Kriging and IDW were compared through CV-assessment and visual examination for five marine pelagic variables measured during the ArcticNet-Malina oceanographic campaign in late summer of 2009. Two variations of IDW were assessed, one in a Euclidean space (IDW_{xyz}) and the other with a restricted vertical search (IDW_{uvw}). Search conditions, IDW-power and variogram range settings (OK) were varied in order to identify an optimal and robust spatial interpolation strategy. For this study, in total, 203 different interpolation strategies were assessed (IDW_{uvw}, IDW_{xyz} and OK, with a combination of method input parameters).

4.6.1 Variogram modeling results

Variogram results indicated some general tendencies. As expected, spatial autocorrelation was stronger in the horizontal than the vertical direction for all variables (Tableau 5), which is a normal consequence of the large variability of marine pelagic variables in the vertical direction as a result of water mass stratification. This vertical anisotropy implied that the ranges of autocorrelation in the vertical direction (direction of lowest continuity) varied between 0.038 km (UHW) and 0.092 km (T), while the ranges in the direction of greatest continuity (horizontal plane) varied between 45 km (T_{subset}) and 276 km (UHW)

With the exception of c_p , a horizontal anisotropy was observed for all variables with a ratio ranging between 1.5 (T_{subset}) and 2.8 (T). However, the azimuth of highest spatial autocorrelation differed between variables: it was shown to be approximately parallel (offset = 7-12 deg.) to Mackenzie shelfbreak for T and UHW while it was in cruise transect sampling azimuth for chl *a* and POC flux (Tableau 5).

The low spatial autocorrelation for POC-flux along shelf break (r^2 , Tableau 5) induced high kriging variances for locations between sampling transects. When assessing the robustness of OK interpolation with an underestimation of variogram derived horizontal ranges, the resulting restricted spatial autocorrelation of 27 km along shelf break was lower than the average distance separating transects. Kriging variances were high between sampling transects and as OK under these conditions could not be justified, interpolation

strategies based on this assessment (OK with underestimation of horizontal ranges) was further excluded from CV analyses. It was consequently concluded that the POC-flux variable was sensitive to variogram modeling. Kriging variances of other variables indicated a general good coverage of the predictor.

Tableau 5. Variogram modeling results. Spherical model used for all variograms.

Variance statistics	Variable					
	T _{subset}	T	UHW	chl a	c _p	POC flux
Nugget	0.00	0.00	0.15	0.05	0.15	0.15
Major horizontal range, r1 (km)	45	276	170	137	76	76
Minor horizontal range, r2 (km)	29	100	95	76	67	41
Vertical range, r3 (km)	0.092	0.18	0.038	0.074	0.080	0.612
Azimuth of major range, r1 (WGS 1984)	88	62	43	2	37	20
Ratio horizontal /vertical range (= r1 / r3)	473	1535	4543	1845	950	125
Geometric anisotropic ratio (= r1 / r2)	1.5	2.8	1.8	1.8	1.1	1.9
Offset angle of r1 to Mackenzie shelfbreak*	33	7	-12	-53	-28	-35

* Offset angle are indicated in clockwise (positive) and anti-clockwise (negative) direction.

4.6.2 CV error results

4.6.2.1 General performance

CV-assessment was carried out in order to identify an *optimal* and *robust* interpolation strategy. As previously mentioned, optimal in this paper refers to a strategy with minimized estimation errors whereas robust refers to a strategy with estimations errors that are stable with respect to a variation of method input parameter. Results of relative method performance and comparison with control (NN interpolation) are presented in Figure 21 a-f, where a lower value of MAE and MSE indicates a better performance of a specific interpolation strategy (full results in supplemental file). Assessments of IDW with vertical restriction of the search neighborhood (IDW_{uvw}) are presented only for UHW, as only minor differences in CV-errors for other variables were obtained (Figure 21).

OrdinaryKriging outperformed the control in terms of error metrics (MAE and MSE) for all variables except c_p (Figure 21 a-f). In the case of c_p, OK produced similar results to the control in terms of MSE, but higher errors with respect to MAE (Figure 21 e). OK performed generally better than IDW methods for T_{subset}, T and UHW (Figure 21 a-c), and with some combinations of search conditions also for chl a, and c_p (Figure 21 d-e). Accordingly, OK produced interpolations with lower or similar errors for all six data sets assessed and was the preferred method. However, compared to the control, IDW-methods yielded lower CV errors for chl a and POC-flux and with optimal combinations of search parameters, also for T_{subset} and UHW Figure 21 a and c).

Influence of search neighborhood

Performance of IDW-methods was stable with regards to a variation of weighted sampling points (*n*) when a power of 5 was used, but not for powers of 1 or 2 (Figure 21 a-e). The lowest CV errors (both metrics: MAE and MSE) were obtained for smaller number of *n* for variables T_{subset}, T and UHW. In other words, with *p* = 1 or *p* = 2, the performance of IDW-methods improved when less importance was attributed to distant sampling

points. An opposite effect was observed for c_p and POC-flux, as MAE diminished with increasing n (MSE was however relatively stable).

OK, including interpolation strategies with an underestimation or an exaggeration of horizontal ranges, was stable with regard to a variation of numbers of weighted samples (n) used in the estimation, for T_{subset} , T , UHW and POC-flux (both metrics). MSE were also stable to a variation of n for chl a and c_p .

4.6.2.2 Robustness

The power (p) of IDW-methods was important for performance. Both MSE and MAE measures were lower when a p of 5 rather than 1 or 2 was used, (all variables except POC-flux; Figure 21). Interpolation strategies with a p -value of 5, were also less sensitive to variations of n . Finally, using a p of 5 also diminished the difference in performance of IDW_{xyz} and IDW_{uvw} (Figure 21 c).

Exaggeration or underestimation of horizontal ranges for OK yielded small modifications of CV errors for T_{subset} , T and UHW. The OK interpolation was less robust with regards to a variation of variogram ranges for other variables. In the case of chl a , an exaggeration of horizontal ranges reduced CV errors (especially in terms of MAE), while an underestimation had the opposite effect. For c_p , both an underestimation and exaggeration of horizontal ranges yielded lower CV errors, which might suggest that variogram model had been misspecified for this variable. As previously mentioned, the interpolation strategy with an underestimation of variogram derived horizontal ranges of POC-flux was discarded due to the high kriging variance obtained.

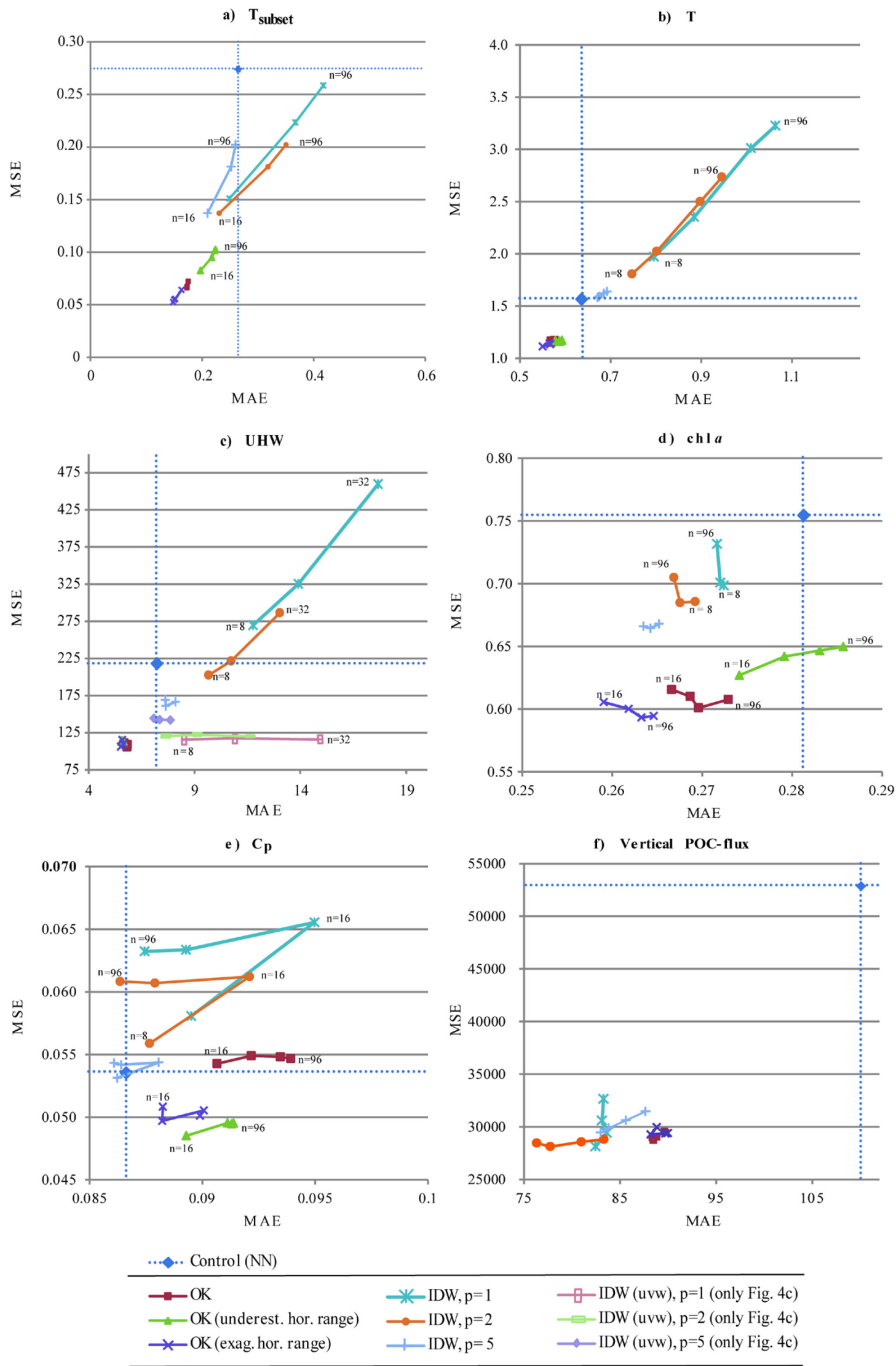


Figure 21. Performance of interpolation strategies (n). MAE (X-axes) and MSE (Y-axes) correspond to mean absolute error and mean square error obtained through CV assessment. IDW with restricted vertical neighborhood (IDW_{uvw}) is illustrated for UHW. n represents the number of neighboring points used in estimations and p refers to the IDW power. Assessment of robustness of OK interpolations was made by an *underestimation* or *exaggeration* of horizontal variogram ranges by a factor of 1.5.

4.6.3 Visual examination of 3D spatial representations

Spatial representations derived from each interpolation strategy (OK, IDW_{xyz}, IDW_{uvw}, and control) with an optimal parameter setting with regards to CV errors were visually compared for each variable. This qualitative comparison of 3D interpolation strategies is presented in this paper by T (Figure 22), that yielded relative high visual differences. As expected, interpolation with NN yielded spatial models with non-oceanographic appearance (distinct cuts) for all variables. Oceanographic phenomena are smoothly distributed, which NN fails to reproduce. The optimal parameter setting for IDW-methods and OK returned spatial models that in general had a sound “oceanographic” appearance. However, IDW-methods sometimes had a tendency to produce circular regions with over-estimations around sampling points, often referred to as ‘bulls eyes’ in the literature (Figure 22 c). This phenomenon was in general less pronounced when a p of 5 was used.

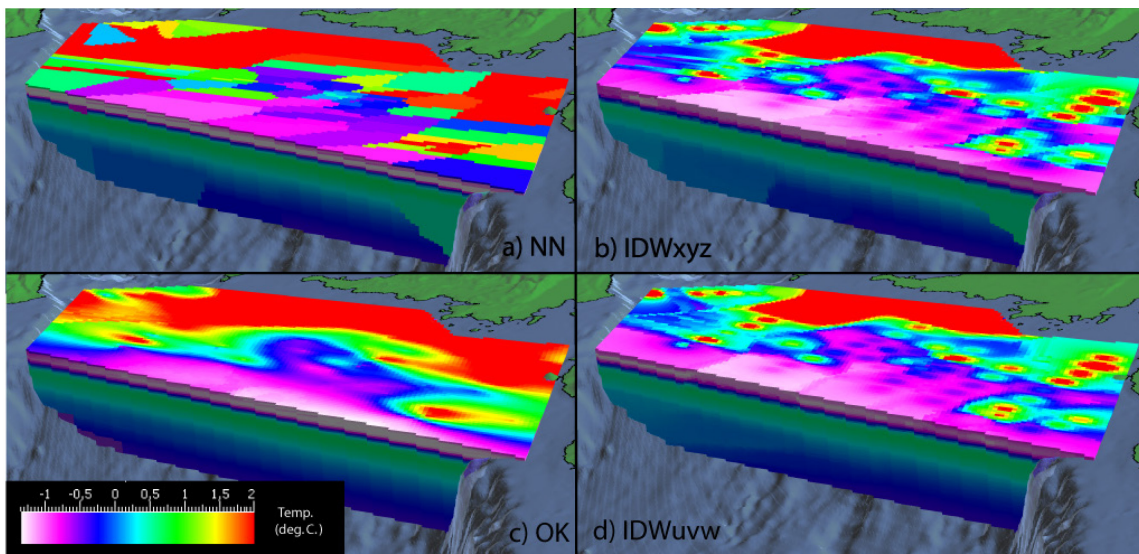


Figure 22. Qualitative comparison of 3D geospatial representations of sea temperature in Beaufort Sea derived from four interpolation method strategies. Range of depth in spatial model is 0 to 1700 m, vertical exaggeration = 50. (a) Natural Neighbor interpolation (control), (b) IDW, no vertical exaggeration of coordinates, $p = 5$, $n = 8$, (c) OK, $n = 32$. (d) IDW in uvw space (exaggeration of vertical coordinates,) $p = 5$, $n = 8$.

4.7 Discussion

4.7.1 3D variogram and spatial autocorrelation

3D variogram modeling was of great value when assessing ranges of spatial autocorrelation of pelagic variables in the Beaufort Sea. All variables studied indicated a presence of spatial autocorrelation in the horizontal as well as in the vertical plane and nugget-to-sill ratios were low, varying from 0 for temperature and 0.05 for chl *a* to 0.15 for UHW, POC-flux and c_p . Accordingly, none of the variables were dominated by a nugget effect. These results indicate that in general terms, 3D spatial modeling with geostatistical methods is appropriate for representation and analyses of pelagic variables.

Somewhat erratic experimental variograms were however obtained in low lag distances for UHW and POC-flux. In the case of UHW, this could be attributed to the relatively low number of data points (290). Only along the shelf break, data density seemed high enough to produce precise theoretical variograms: irregularities for low-distance bins in other azimuths were assumed to be an artifact of this low number of samples. An insufficient number of sampling stations was also limiting for POC-flux, as this variable showed a low range of spatial autocorrelation in the horizontal plane. Notably, recent work by Forest et al. (2013) indicates that POC-flux on the Mackenzie shelf is distributed in a patchy manner and that local variation is important. Nevertheless, the relatively good performance of OK compared to control and IDW-methods justifies the use of 3D variogram modeling, even when sampling is sparse.

4.7.2 Method performance

4.7.2.1 Overall method performance

Comparing performance of interpolation methods in this study to previous research is difficult, as no other comparative study has been conducted, to our knowledge, that assesses 3D interpolation methods in the marine pelagic environment. Several studies exist that explore 3D interpolation performance in a geological context (Falivene, Cabrera, et Saez 2007; Falivene et al. 2010; Jones et al. 2009) or for modeling of soil texture (He et al. 2009; He, Hu, Chen, et al. 2010), but these studies are mostly limited to interpolation of categorical variables. In a review of 18 comparative studies performed in 2D environment (including meteorology, water resources, ecology, agricultural and soil science as well as marine environmental science), Li and Heap (2011) found that published papers presented a 'dramatic variation' in terms of method performance. OK and IDW were the methods figuring in the largest amount of comparative assessments. Their average performance in terms of *relative* MAE (MAE / mean of data set) was around 15 %, which is comparable to the CV-errors obtained for the optimal search strategy in this research for T_{subset} and c_p (Figure 23), despite the higher dimension (3D) of our study.

It is well acknowledged that accuracy of predictions decrease for variables with increased data variance (Kerry et Oliver 2007; Li et Heap 2011; Olea 2007; Park et al. 2012). This is consistent with our results as T_{subset} and c_p were characterized by a relative low data variance, whereas higher method errors were obtained for chl *a* and vertical POC-flux, with greater data variance (Figure 23).

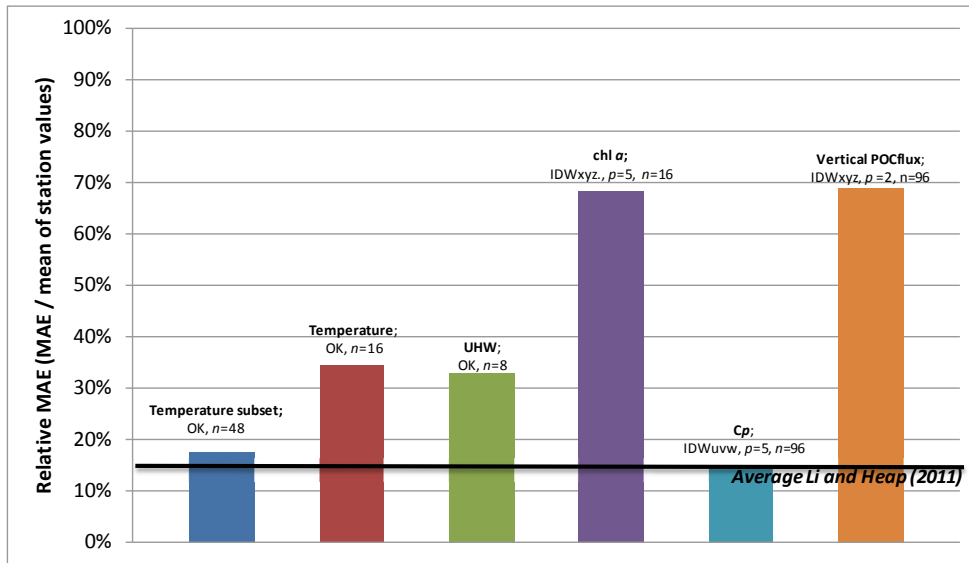


Figure 23. Relative MAE of interpolation strategies with an optimal searching neighborhood (n). Y-axis represents the *relative MAE* (MAE / arithmetic mean of variable) and vertical bar represents the approximate average of both OK and IDW performance in a comparative study of 2D CV-assessments conducted by Li and Heap (2011).

4.7.2.2 Optimal interpolation method

OK largely outperformed IDW-methods for T_{subset} , T and UHW that all corresponded to variables with a relative high spatial autocorrelation in the Mackenzie shelf break azimuth. The weaker spatial autocorrelation in this azimuth for other variables (especially POC-flux), imposed a restriction of the OK interpolation method as the resulting search neighborhood only permitted limited consideration of sampling points from adjacent sampling transect. This restriction potentially adversely affected the performance of OK for these variables. Nevertheless, CV-error assessment indicated that OK was still the best choice for spatial interpolation of chl a , and that no clear difference in performance between IDW-methods and OK was observed for c_p or POC-flux. CV-errors for c_p were in general low (Figures 21 and 23), suggesting that either of the three interpolation methods assessed (OK, IDW $_{xyz}$, IDW $_{uvw}$) could be used to accurately interpolate this variable as long as method input parameters were optimized.

4.7.2.3 Assessing IDW with vertical restriction of search neighborhood

Only for UHW, a notable difference in CV-errors was obtained by using IDW with a vertical restriction of search neighborhood (IDW $_{uvw}$). This was somewhat surprising considering the vertical stratification of variables. Visual examination of final spatial models also confirmed that the two variations of IDW yielded comparable qualitative results. In our study, the vertical exaggeration was deduced from grid design and not from variogram modeling as in one recent study (Falivene, Cabrera, et Saez 2007). The use of a variogram derived

ratio between horizontal and vertical ranges to constrain vertical search neighborhood would potentially have led to an increased difference between IDW_{xyz} and IDW_{uvw} . However, this approach requires going through 3D variogram modeling steps and also presumes accuracy of 3D variograms. Major advantages of IDW method (simplicity, less time-consuming) are then lost.

4.7.2.4 *Influence of search conditions*

Search conditions (n), have previously been recognized as important for the performance of IDW in the context of spatial modeling in 2D (Babak et Deutsch 2009; Kravchenko et Bullock 1999; Weber et Englund 1994; Zimmerman et al. 1999) and 3D (Falivene, Cabrera, et Saez 2007). This was confirmed for 3D in this research, especially when a power of 1 or 2 was used. With a power of 5, most of the IDW weight is attributed to the closest observation and estimations are relatively stable with respect to a change of n . OK performance was relatively constant to a variation of n , especially with regard to MSE. This is also consistent with previous research performed in a 2D spatial modeling context (Weber et Englund 1994).

When the nugget-to-sill-ratio is low and nears zero (between 0 and 15% of total sill in this study, Tableau 5) and the range is relatively important compared to the study zone, the closest samples are highly correlated. In this case, only nearby samples could be required to ensure an accurate estimation (Vann, Jackson, et Bertoli 2003) and the number of n has little effect on performance of estimations. The principal impact of increasing n is then an increase in the smoothing effect of kriging. No relation between nugget-to-sill ratios and optimal number of n was observed in this study, which is probably explained by the overall low nuggets. It has been proposed that a higher number of samples (n) should be used for variables with high data variance in order to capture the spatial variation (Li et Heap 2011). No clear tendency of this kind was supported by our research. The optimal number of n could not be related neither to the number of sampling points used in data sets (varying between 290 for UHW and 9369 for T_{subset}), nor to the vertical distance between sampling points (1 m for T_{subset} and 70 m for UHW) or the vertical range of spatial autocorrelation (38 m for UHW and 180 m for T).

4.7.2.5 *IDW and OK robustness*

Our results confirm that IDW in 3D spatial modeling is sensitive to the power, which is also consistent with previous studies in 2D (Babak et Deutsch 2009; Li et al. 2008; Li et al. 2010; Weber et Englund 1992). While a value of 2 is most commonly used and a default value in many GIS, this value is not always the most appropriate. In a 2D interpolation of marine salinity and temperature a p of 4 was found to be most appropriate (p of 1, 2 and 4 assessed; Li et al. 2010). In general terms (2D assessments), it has been noted that when data points are scarce or missing, increasing the exponent used in IDW can improve the performance of interpolation results (Li et al. 2008). Babak and Deutsch (2009) found that a higher p (4 and 6) was preferable

to low values (1 - 3) when n increased. In our study, a p of 5 was the best choice for all variables assessed as CV error decreased and method robustness (regarding variation of n) increased.

OK robustness was assessed by varying the horizontal ranges of spatial autocorrelation derived from variogram modeling. In general terms, an exaggeration of horizontal ranges did not decrease OK performance. In the case of chl a , results indicated a slightly better performance when exaggerated horizontal ranges of autocorrelation were used. This indicates that variogram model was not accurate. The highest spatial autocorrelation was identified along sampling transect azimuth for chl a (and to a certain degree also for c_p and vertical POC-flux; Tableau 5). This might first be surprising, as spatial autocorrelation was assumed to be stronger parallel to coast line and shelf break. However, the ability to estimate variogram parameters is affected significantly by the spatial configuration of sampling points (Muller et Zimmerman 1999). As sample density was higher along azimuth parallel to sample transects, we can assume a more precise estimation of spatial ranges for these azimuths. The higher range in this azimuth would then (at least partially) be an artifact of an insufficient sampling in shelf break azimuth. Considering this dilemma, it has been suggested previously that an exaggeration of horizontal ranges in azimuths of lowest sample density could be a general strategy to improve OK performance. Another explanation of the higher spatial autocorrelation in cruise sample direction of chl a could be a temporal evolution of chl a due to the timing/zone of sampling: time lag was more important *between* sampling transects than *within* transects.

4.7.3 Clustered sampling and temporal variation influences

4.7.3.1 Data clusters and CV assessment limitations

It has been acknowledged that performance of spatial interpolation methods decrease for clustered data sets compared to regularly distributed data (Zimmerman et al. 1999). The ArcticNet-Malina data set used in this research was derived from a sampling design typically found in oceanographic research where stations are aligned along transects perpendicular to coastline. No transformation of data was performed in order to account for this clustering as it was considered to be a part of this research to investigate the performance of 3D spatial modeling techniques in this particular sampling context. It is important to notice that our data sets contained no preferential sampling for areas of higher or lower values. In other words, clustering was rather limited to the configuration of sampling points. Such clustering will, at least partly, be accounted for by OK, as closely spaced data will inform fewer grid nodes and receive lesser weight compared to widely spaced data (Deutsch 2002). In addition, spatial clustering was low for the reduced T_{subset} and UWW data sets and method performance of these variables could thus be compared to other variables assessed in order to discuss the effect of clustering. Nevertheless, the string-effect of kriging (Deutsch 1994) was not accounted for in this study which might have decreased the performance of OK compared to IDW. Daly (2006) points out that

spatially clustered data sets may yield an underestimation of CV-errors because such spatial configuration favors the possibility that a nearby station will be present to produce a good estimate for the station tested. No such evident trend was observed in this research, as the lowest CV-errors were observed for c_p (clustered data set) and T_{subset} (non clustered data set, Figure 23). We also note that average CV-errors for T (full spatial coverage) were twice as high as for T_{subset} (reduced spatial coverage). The spatial clustering of data along transects perpendicular to shelf break then seems to have been responsible for an increase in CV-errors for this variable. An alternative explanation is that the higher data variability for T (-1.44 - 8.38 °C compared to T_{subset} (-1.44 - 1.99 °C), or the prolonged sampling period (38 days compared to 10 days), diminished performance of spatial interpolations methods for T compared to T_{subset} . According to Weber and Englund (1992) kriging would be favored compared to IDW in circumstances where biased clustering of sampling exists.

4.7.4 Final volumetric models and ecological importance

In this research, we focused on the potential and performance of 3D spatial interpolation and modeling methods of marine pelagic variables. An in-depth analyses of the spatial variation of pelagic variables treated in this research as well as the physico-chemical conditions that prevailed in the south-eastern Beaufort Sea during the ArcticNet-Malina campaign is out of the scope of this study and the reader is referred to recent studies (Forest et al. 2013). However, a short discussion of results in an ecological context is important primarily for model validation purposes, but also with the intention to show some advantages of 3D spatial modeling. Note that a rigorous investigation of scale and geographic position in following volumetric models (e.g., depth of a certain phenomenon.) requires interactive visualization tools, available for figures in their original form (Paradigm GOCAD® version), but not through this paper.

As expected, hot-spots of temperature maxima are found primarily in inshore regions (depth < 100 m) and values above 1°C (Figure 24) are observed in surface layer (0-36 m depth) all over shelf area. Temperature maxima were also observed close to the Mackenzie delta and offshore to the Mackenzie canyon. The homogeneous stratification of temperature correlates with the presence of distinct water masses and is thus well correlated with the spatial distribution of UHW. At an approximate depth of 225 m, a distinct stratification (blue colored layer, Figure 24) represents the fuzzy boundary between UHW and Atlantic water.

The 3D spatial model of UHW indicates a relatively uniform stratified distribution with maxima situated at a depth between 110 and 190 m (Figure 25). Within this range, UHW constitutes more than 80 % of water volume. This is consistent with earlier observations on the Mackenzie shelf (2003 – 2004), confirming off-shelf maxima at water depth varying between 120 and 180 m (Lansard et al. 2012). The UHW layer reaches some deeper depths off the Mackenzie Canyon and at the entry of the Amundsen Gulf (max ~250 m, both sites).

The 3D model also indicates a relatively high fraction (>50 %) of UHW in the Mackenzie Canyon at an approximate depth of 50 m, which is consistent with the presence of enhanced shelf break upwelling in this area (Lansard et al. 2012). Such events can inject new nutrients from the sub-surface depth into the nutrient-poor surface layer and typically constitute hot-spots for biological activity (Forest et al. 2007; Sampei et al. 2011; Tremblay et al. 2011). In summer 2009, only moderate upwelling was observed nearby Cape Bathurst, located at the entrance to Amundsen Gulf, an area that has previously been recognized as a major spot for upwelling (Sampei et al. 2011; Williams et Carmack 2008). However, most of the Malina campaign (later than early August) was in general characterized by the absence of upwelling-favorable wind as observed in previous campaigns (Forest et al. 2013).

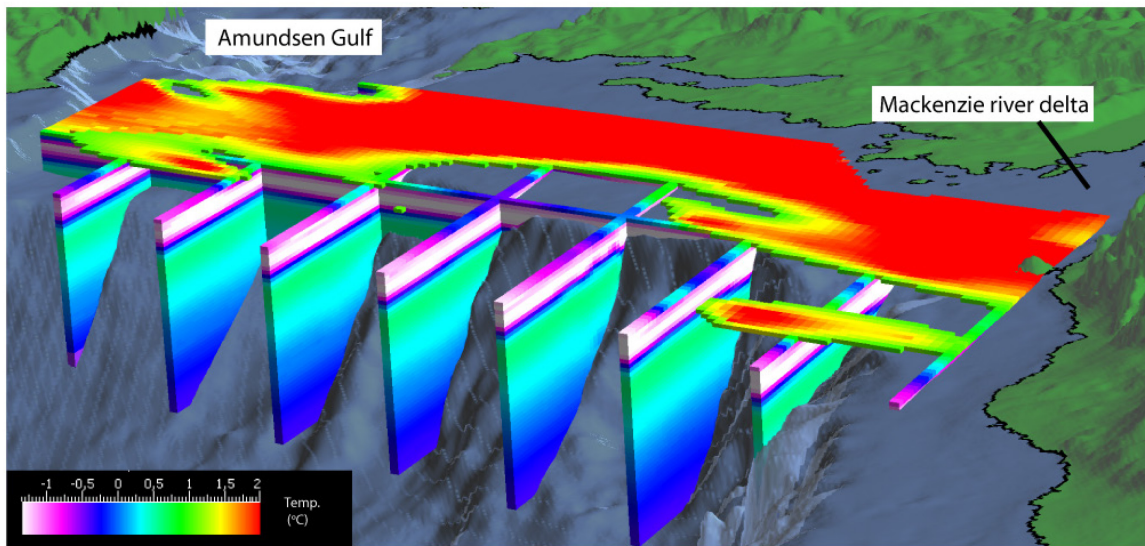


Figure 24. Volumetric model of temperature, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. 3D interpolation strategy: OK, $n = 32$. Depth represented of spatial model is 0 to 1700 m. Volume represents grid cells where temperature is superior to 1°C. Vertical exaggeration = 50 in figure.

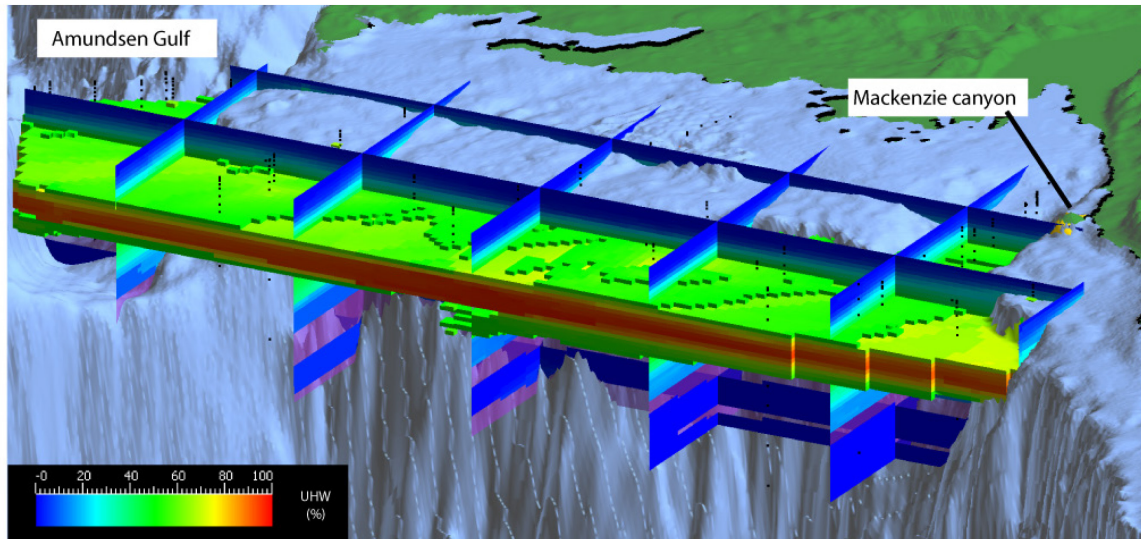


Figure 25. Volumetric model of UHW, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. Fraction of water mass constituted of UHW (%) for the upper 500 m of water column. Interpolation: OK, $n=32$. Volume represents voxels where UHW fraction is $>50\%$. Black dots indicate sampling points. Areas with no-data values are colored transparent mauve. Vertical exaggeration below sea-surface = 200.

Identification of geographical chl *a* maxima can be used for identification of biological hot-spots as chl *a* concentration is a common metric for phytoplankton biomass, and thus for primary production occurring in the oceans (Martin et al. 2002). However, recent studies indicate that c_p could be an alternative metrics for phytoplankton and more adequate for open ocean environments, provided that it actually reflects phytoplankton particulate organic carbon (Behrenfeld et Boss 2003, 2006). Highlighting similarities and differences of variable maxima of these variables can then be of value in the discussion of appropriate metrics for phytoplankton biomass and biological hot-spots. Our results reveal a clear overlap for both chl *a* and c_p hot-spots (Figure 26). These are primarily concentrated to areas previously recognized with relative importance of biological production, such as upwelling areas in the proximity of shelf break, Cape Bathurst and in the vicinity of the Mackenzie canyon. The region with relatively high values of c_p observed close to Mackenzie river delta could be explained by the presence of allochthonous Mackenzie river run-off particles that would not directly induce enhanced levels of chl *a*. A SCM was centered offshore to an approximate depth of 60 to 70 m, which is consistent with other reported analyses from the ArcticNet-Malina campaign (Forest et al. 2013). The importance of SCM for the total productivity of the Arctic Ocean has recently been emphasized (Ardyna et al. 2013; Martin et al. 2010). In comparison, our 3D spatial models indicate that offshore maxima of c_p do not follow any clear spatial trend, presumably as a result of intermediate nepheloid layers propagating from the shelf break (Forest et al. 2007).

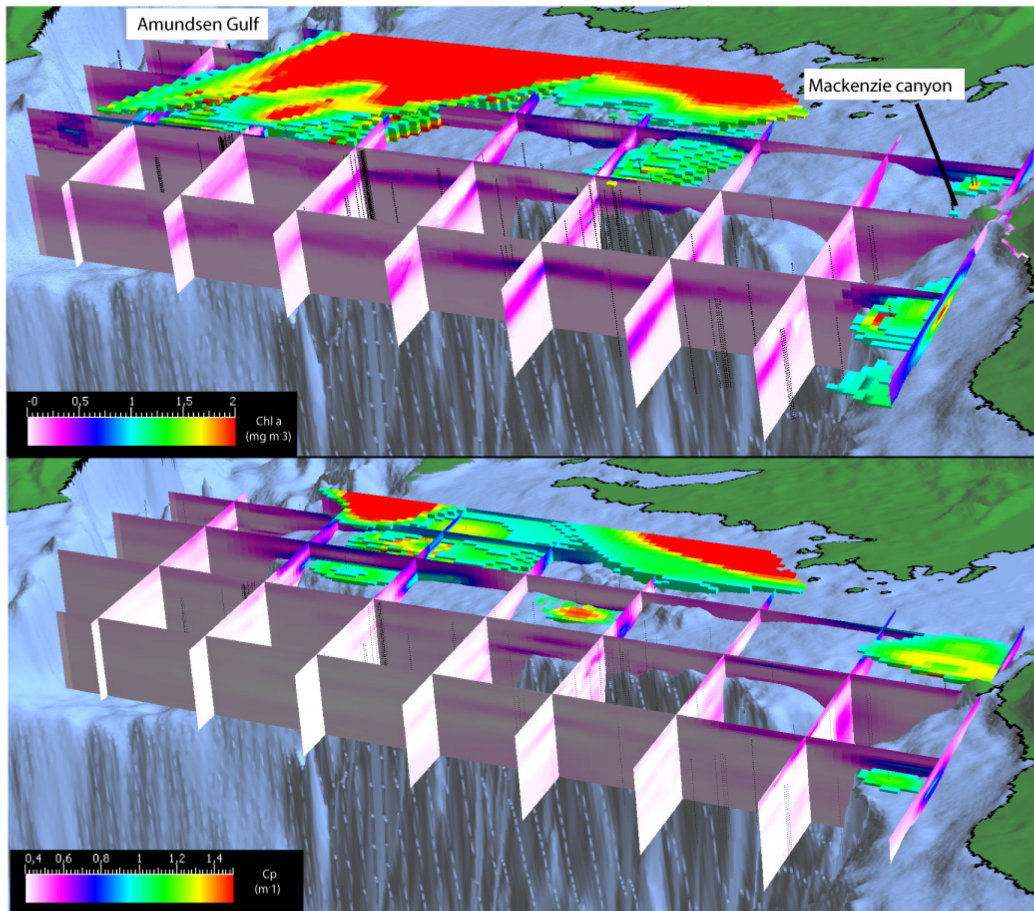


Figure 26. Volumetric model of chl *a* et c_p , during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. Spatial models represent 200 upper meters of water-column. Black dots indicate sampling points. Vertical exaggeration = 200. (a) Chl *a*, interpolation strategy = OK, $n = 32$. Volume for voxels illustrated when chl *a* > 1 mg m⁻³ (b) C_p , interpolation strategy = IDWxyz, $\rho = 5$, $n = 48$. Volume for voxels illustrated when c_p > 1 m⁻¹.

With the objective to determine the role of the Arctic Ocean in the global oceanic carbon cycle, it is crucial to estimate total primary production and gain understanding of biological hot-spots, but also to elucidate net transport of carbon towards the deeper part of the ocean. In this context, the study of vertical POC-fluxes is essential as a fraction of POC-flux produced in surface layers is retained and recycled in biological local or regional cycles (Forest et al. 2007; Sampei et al. 2011). Our 3D spatial model of POC (Figure 27) permits a flexibility of vertical and horizontal cutting planes in any azimuth that can give a better insight to spatial variations of POC-fluxes than traditional vertical cutting planes in a 2D environment limited to transect azimuths. Indeed, major features of spatial variations noted in a traditional 2D environment (Forest et al 2013) are also represented by our 3D spatial model. Additionally, the 3D spatial model allows considering irregular sample points (ArcticNet campaign) that can be included in analyses. POC-fluxes are highest on the shelf with

hot-spots close to Cape Bathurst and in the Mackenzie canyon. We also notice the patchy distribution of POC fluxes close to shelf break (center of the study area), as confirmed by recent research (Forest et al. 2013).

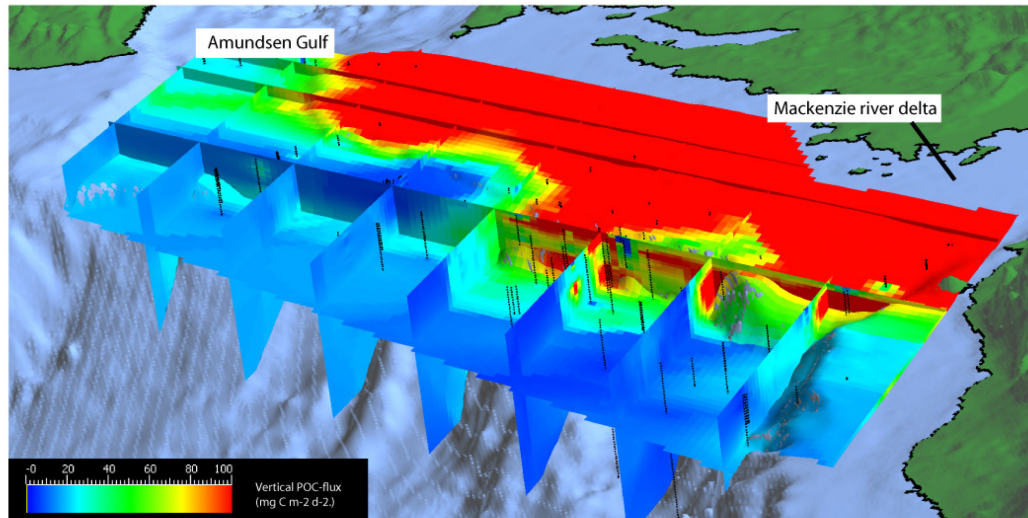


Figure 27. Volumetric model of vertical POC-fluxes, during ArcticNet-Malina campaign (south-eastern Beaufort Sea) in 2009. 3D interpolation strategy: IDW_{xyz} , $p = 2$, $n = 48$. Upper 1700 m of water-column illustrated. Horizontal cuts are situated at 600 m depth in offshore areas and at 110 m in shelf area (or at bottom for depths <110 m). Black dots indicate sampling points. Vertical exaggeration = 50.

4.8 Conclusions

In this paper, we assessed the performance of 3D spatial interpolation and modeling for five marine pelagic variables with increasing complexity regarding spatial autocorrelation, sampling heterogeneity and ecological function. Performance of Ordinary Kriging: OK (geostatistical method) was compared to Inverse Distance Weighting: IDW (deterministic method) with the objective to identify an optimal and robust interpolation strategy. The results and general recommended approach for 3D interpolation of marine pelagic environment can be summarized as follows:

1. 3D variogram modeling is useful in predicting spatial autocorrelation of marine pelagic variables.
2. Overall, OK yielded either lower (T_{subset} , T and UHW) or similar (chl a , c_p POC-flux) CV-errors than IDW and was hence the optimal method for pelagic variables assessed. However, when complexity of pelagic variables increased in terms of spatial autocorrelation and data variation (chl a , vertical POC-flux), difference between OK and IDW in terms of CV-errors diminished.
3. A lower number of n (~8-16) and a higher number of p can increase performance and stability of IDW in the marine pelagic environment.

4. OK was the more robust interpolation method as it was generally stable to variations of method parameters, while IDW was sensible both to a variation of numbers of neighbors (n) and the weight (ρ) attributed to neighboring points in the estimation.
5. General performance of our 3D spatial interpolations was comparable to that noted in 2D for c_p , T, UHW, but diminished for variables with a higher data variance or lower spatial autocorrelation (chl a , vertical POC-fluxes).

Accordingly, if data sampling is sufficient to provide reliable 3D variograms, OK is preferred to IDW for spatial interpolation of the marine pelagic environment. Geostatistical interpolation techniques are more time consuming, but at the same time, the use of IDW methods needs, to a higher degree, be supported by an assessment of method input parameters. In the specific case of the ArcticNet-Malina campaign, an increased performance of spatial interpolations would have been obtained with a less clustered sampling design.

Our results show that recent advances in volumetric representation and 3D interpolation techniques can be used to extend usual interpretation of static marine pelagic phenomena from static planes derived from a 2D to a 3D environment. Indeed, we consider that such an upgrade can enhance the study of the marine pelagic environment. 3D spatial modeling of the marine pelagic environment is promising, which is illustrated in this paper by volumetric raster models constructed with the geomodeling tool Paradigm GOCAD® (version 2009.2). However, geomodeling tools could be further adapted to the marine environment by improving basic functions such as image treatment, map algebra and spatial data base management. At the same time, integration of volumetric representation in a GIS environment would be an important advance towards an optimal marine GIS. This tool must also include representation, interpolation and analyses functions of 2D static vertical cutting planes, absent from commercial GIS like ArcGIS10. In addition, all spatial modeling tools conceived for the pelagic environment would benefit from including spatial data structures that permits to handle the temporal (dynamic) characteristics of the pelagic environment.

This work represents a first attempt to highlight advantages of recent development in 3D geostatistics and geomodeling for use in the study of the pelagic ecosystem. Within the scope of this paper, we limited our work to an assessment of two of the most commonly applied interpolation methods. Many other are further to be assessed and some would probably increase the precision of our estimations. Block-estimation (or block kriging) represents one of those having a high potential for pelagic variables. In further works, a refining of interpolation methods and especially of the kriging search strategy will be considered, which has the potential to further increase overall estimation accuracy of OK. Such a refining will also include an assessment of different declustering methods and techniques that support parameter setting, such as quantitative kriging neighborhood analysis.

4.9 Acknowledgements

The authors would like to thank Bruno Lansard who shared data on Upper Haloclinic Water. We also express gratitude to L. Prieur, M. Picheral, C. Marec and Y. Gratton for the deployment of the CTD/UVP-rosette and for the processing of data, as well as P. Guillot for the validation of CTD data. Finally, we are grateful for the financial contribution to this research from Natural Sciences and Engineering Research Council of Canada.

CHAPITRE 5 - Conclusion générale

5.1 Contribution de la recherche

L'apport principal de cette étude est d'avoir démontré le potentiel et la pertinence de la modélisation 3D pour l'environnement pélagique marin. Nous avons confirmé que les outils émergents de visualisation scientifique dans le domaine de la modélisation géologique peuvent servir à améliorer l'étude de cet écosystème. Les performances de la modélisation spatiale 3D pour des jeux de données issus d'une campagne océanographique typique sont satisfaisantes, malgré les particularités de l'environnement marin à l'égard de la variabilité spatio-temporelle des variables et de la répartition typique des points d'échantillonnage dans l'espace.

Cette étude a également permis de combler une lacune importante identifiée dans la littérature scientifique, en examinant – pour la première fois, à notre connaissance – les performances des méthodes d'interpolation 3D des variables continues dans l'écosystème pélagique marin. Un nombre élevé de publications compare les performances des méthodes d'interpolation en 2D, mais peu d'études traitent des performances de la géostatistique et des méthodes déterministes pour les variables continues des champs volumétriques. Il s'agit donc d'un pas important vers le développement de méthodes spécifiquement conçues pour la modélisation spatiale 3D de l'écosystème pélagique marin.

Les résultats de cette recherche ont été présentés plusieurs fois à l'équipe d'océanographes du laboratoire Takuvik, et au public lors de la conférence « 7th 3DGeoInfo » qui s'est tenue à Québec en mai 2012. Une partie des résultats a été publiée dans deux articles scientifiques, le premier sous la forme d'un article court pour les annales de la conférence 3DGeoInfo (Sahlin, Mostafavi et al. 2012), et le second sous la forme de matériel complémentaire dans un article océanographique (Forest, Babin et al. 2013). Le cœur de cette étude a fait l'objet d'un troisième article, actuellement évalué par le comité de lecture de la revue scientifique *Marine Geodesy*.

5.2 Atteinte des objectifs

L'objectif général de ce projet de recherche consistait à vérifier si la modélisation spatiale 3D (représentation et analyse) de l'écosystème pélagique marin pouvait aider à la compréhension de son fonctionnement. Cet objectif a été atteint à travers trois sous-objectifs.

Dans un premier temps, nous avons identifié les difficultés propres à la modélisation spatiale 3D de l'écosystème pélagique et évalué le potentiel des nouveaux outils. L'analyse de la littérature a permis de constater l'absence d'outil adéquat et bien adapté pour la modélisation spatiale 3D de l'écosystème pélagique.

Les efforts récemment consentis pour le développement des SIG 3D ont mis l'accent sur la vue conceptuelle orientée-objet en utilisant des structures vectorielles, ce qui ne constitue pas une solution adéquate pour la modélisation des champs volumétriques. À cet égard, les outils de modélisation géologique ont un grand pas d'avance sur les SIG, et leur potentiel pour la modélisation de l'écosystème pélagique a été démontré dans notre étude. Cependant, les avantages typiques des SIG (p.ex. : transformation de coordonnées, gestion et analyse de bases de données, analyse spatiale) sont absents ou limités dans les outils de modélisation géologique; une meilleure adaptation de ces derniers aux contraintes particulières de l'environnement marin est donc nécessaire.

Dans un deuxième temps, nous avons validé les performances de l'interpolation spatiale 3D pour cinq variables d'une complexité croissante par rapport à l'anisotropie des jeux de données et leur fonction écologique. Un outil de modélisation spatiale identifié dans la première phase du projet, Paradigm Gocad® a été utilisé pour cette évaluation. Les performances et la précision des méthodes d'interpolation 3D ont été vérifiées sur la base de métriques d'erreur, ainsi que par l'examen visuel des modèles spatiaux finaux.

Les résultats de la *validation croisée* (CV) indiquent que la précision obtenue est généralement comparable à celle observée pour l'interpolation spatiale 2D pour les variables de T. (sous-ensemble ou non), C_p , et UHW. Les performances des interpolations spatiales 3D pour chl *a* et le flux vertical de POC étaient toutefois plus faibles, ce qui s'explique par plusieurs facteurs : (1) la complexité croissante de la variabilité spatio-temporelle de ces paramètres avec des maxima dans des stations côtières entre 80 (flux COP) et 150 (chl *a*) fois les valeurs médianes; (2) une approche conservatrice de CV impliquant plus de poids pour les points de mesures dans les stations côtières; (3) une autocorrélation spatiale généralement plus faible que pour les autres variables. Malgré les erreurs plus importantes pour chl *a* et le flux de COP, les modèles spatiaux obtenus sont d'un niveau de détail satisfaisant comparé à ceux obtenus avec d'autres techniques d'analyse océanographique pour les mêmes jeux de données (Forest, Babin et al. 2013) ou des jeux de données antérieures dans la même région. On pourrait ainsi conclure que les erreurs associées à l'interpolation spatiale 3D pour l'ensemble des variables sont acceptables dans ces conditions et que la modélisation 3D peut être utilisée pour améliorer l'étude de l'écosystème pélagique marin, ce qui indique l'atteinte du deuxième sous-objectif de la recherche.

Ce sous-objectif visait également à comparer les performances des stratégies d'interpolation 3D géostatistiques et déterministes pour l'écosystème pélagique marin et à identifier une approche d'interpolation optimale et robuste. Deux des méthodes d'interpolation les plus couramment utilisées pour la modélisation spatiale 2D ont été comparées dans ce contexte : la méthode déterministe de distance inversée (IDW) et la

méthode géostatistique de krigeage ordinaire (ordinary kriging : OK). Une variation de la IDW avec une exagération de la coordonnée verticale faisait également partie de l'analyse.

Des études antérieures en 2D (Li and Heap 2008; Li and Heap 2011; Li, Heap et al. 2011) ont conclu que les performances de l'interpolation spatiale dépendent de la méthode choisie, des paramètres spécifiés pour chaque méthode, de la répartition des points de mesure dans l'espace et de la variance des variables. Les résultats de cette étude indiquent que si un variogramme 3D fiable peut être construit, la méthode OK fonctionne mieux que les deux méthodes IDW pour l'interpolation spatiale de l'environnement marin pélagique. On note également que :

- ✓ OK demeure généralement stable malgré les variations dans les paramètres de la méthode, alors qu'IDW est sensible à la fois à une variation du nombre de voisins (n) et au poids (p) attribué à des points voisins dans l'estimation;
- ✓ pour les approches IDW, une valeur faible de n (~ 8-16) et une valeur plus grande de p (5 dans cette recherche) peuvent augmenter les performances;
- ✓ lorsque la complexité spatio-temporelle des variables pélagiques augmente en termes d'autocorrélation spatiale et de variance (chl a , flux COP), la différence entre la performance d'OK par rapport aux approches IDW diminue.

Finalement, un dernier sous-objectif dans le cadre de cette étude était de proposer des représentations spatiales 3D pour des variables pélagiques issues de la campagne océanographique ArcticNet-Malina. On a déjà démontré que la précision obtenue par l'interpolation spatiale 3D comparée à des études antérieures était acceptable (sous-objectif 1), et qu'OK était à préférer lorsqu'il était possible d'obtenir un variogramme 3D fiable (sous-objectif 3). Les représentations optimales en termes de minimisation des erreurs MAE et MSE ont été présentées dans le chapitre 4.

5.3 Perspectives et futurs travaux

5.3.1 Développement d'un SIG marin 3D

Le marché actuel ne semble pas offrir d'outils de modélisation spatiale 3D bien adaptés à l'étude de l'écosystème pélagique. La revue des SIG communs réalisée dans le cadre de cette recherche indique que ces logiciels n'ont pas encore les fonctions nécessaires à l'interpolation, la représentation et la visualisation 3D

des phénomènes continus : trois éléments indispensables pour l'étude spatiale 3D des écosystèmes marins. L'intégration de ces fonctionnalités serait une avancée importante vers un SIG marin 3D optimal. Cet outil doit aussi inclure des fonctionnalités de représentation et d'analyse de coupes statiques verticales ou horizontales dans un environnement 2D. Les performances des outils de modélisation géologique pour la modélisation spatiale 3D de l'environnement pélagique marin sont prometteuses, malgré l'absence (ou la limite) des avantages typiques des SIG, comme la gestion et l'analyse de bases de données. Les fonctionnalités de gestion et d'analyse spatiale des couches raster 2D sont pratiquement absentes d'un outil comme Paradigm GOCAD® (ver. 2009.2), et pourtant essentielles puisque l'océanographie repose partiellement sur la télédétection (p. ex. images satellite) pour la prise de données. En outre, l'évaluation des performances de l'interpolation spatiale 3D (effectuée avec Paradigm GOCAD®) aurait été grandement facilitée si des fonctions de mesure d'erreur (CV) automatisées avaient été disponibles dans les outils comme ArcGIS (outils géostatistiques, en 2D et 2,5D).

De façon générale, on note que la distinction entre les outils de modélisation géologique et les SIG s'efface tranquillement, alors que dans chaque domaine on emprunte des fonctionnalités à l'autre. Les outils géostatistiques ArcGIS sont un exemple en ce sens. Le développement des *plug-ins* commerciaux dans les SIG, comme *EnterVol for ArcGIS*, est un autre exemple. EnterVol a été conçu pour étendre la modélisation spatiale des phénomènes continus à un environnement 3D dans ArcGIS, et a été développé par une société spécialisée dans le domaine de la modélisation géologique (CTech). De tels *plug-ins* indiquent possiblement la mise en marché prochaine d'une première génération de SIG marins 3D pour la modélisation volumétrique (notons toutefois que les fonctionnalités géostatistiques d'EnterVol sont encore beaucoup trop restreintes pour permettre une démarche de modélisation spatiale rigoureuse).

5.3.2 Modélisation et interpolation spatiale 3D pour l'écosystème pélagique marin

Le développement de la modélisation numérique, destinée à prédire la dynamique de l'océan et les processus biogéochimiques, est au cœur de l'étude de l'écosystème pélagique. Beaucoup moins d'efforts sont dédiés à la modélisation spatiale de cet environnement. Pourtant, la représentation spatiale des phénomènes continus de l'écosystème pélagique constitue l'une des méthodes principales pour l'analyse et la représentation de cet environnement (p.ex. Forest, Babin et al. 2013, van Ruth, Ganf et al. 2010, Martin, Tremblay et al. 2010).

Dans ce mémoire, deux des méthodes d'interpolation les plus couramment utilisées pour la modélisation spatiale 2D ont été comparées dans le contexte 3D, soit OK et IDW. Toutefois, beaucoup d'autres méthodes d'interpolation spatiales existent et certaines d'entre elles auraient potentiellement amélioré la performance de l'interpolation et donc la représentation des variables pélagiques. Dans l'optique de poursuivre ces travaux d'évaluation, plusieurs méthodes d'interpolation sont pertinentes, dont le krigeage universel et le cokrigeage

ordinaire. Alors que le premier permet d'améliorer l'interpolation lors de la présence de tendances spatiales dans un jeu de données (p.ex. la stratification à la verticale), le deuxième permet de considérer la distribution spatiale d'une variable secondaire dans l'interpolation. Finalement, la performance des méthodes d'interpolation spatiale 3D dans cette recherche a été effectuée dans deux espaces de raster 3D distincts, soit un premier où les voxels le long de l'axe Z étaient parallèles à la surface de l'eau et un deuxième où les voxels étaient comprimés entre la surface et le fond (section 2.2.2.1). Un espace de modélisation avec des voxels de taille irrégulière qui suit les gradients de densité de l'eau aurait possiblement amélioré l'interpolation pour certaines des variables. Par contre, la définition de ces gradients à travers l'espace de modélisation aurait aussi nécessité une interpolation spatiale.

Il est à souligner que la répartition de l'échantillonnage le long de radiales perpendiculaires à la côte (comme pour Malina) est responsable d'une agrégation des données qui diminue la performance des méthodes d'interpolation. Cette diminution semble notamment être attribuable à la difficulté de modélisation du variogramme, mais également à la méthode de validation (CV) elle-même. Comme le démontre cette recherche, en considérant le potentiel de la modélisation spatiale 3D pour l'étude des variables continues de l'écosystème pélagique, nous suggérons que les plans d'échantillonnage, soient optimisés a priori pour permettre une meilleure distribution spatiale des stations d'échantillonnage. Une répartition irrégulière des stations permettrait en effet d'augmenter le détail de la représentation spatiale volumétrique et donc la compréhension de l'écosystème. Ce changement d'habitude viendra en son temps, mais j'espère avec ce mémoire avoir contribué à cette perspective de développement.

Bibliographie

- Altman, D. 1994. Fuzzy set theoretic approaches for handling imprecision in spatial analysis. *International Journal of Geographical Information Systems* 8 (3):271-289.
- Angel, M. V. 1993. Biodiversity of the pelagic ocean. *Conservation Biology* 7 (4):760-772.
- ArcticNet. 2010. Coming together in the study of a changing Canadian Arctic.: Université Laval.
- Ardyna, M., M. Babin, M. Gosselin, E. Devred, S. Bélanger, A. Matsuoka, et J.-É. Tremblay. 2013. Parameterization of vertical chlorophyll a in the Arctic Ocean: impact of the subsurface chlorophyll maximum on regional, seasonal and annual primary production estimates. *Biogeosciences Discussions* 10:1345-1399.
- Arsenault, R., C. Ware, M. Plumlee, S. Martin, L. L. Whitcomb, D. Wiley, T. Gross, et A. Bilgili. 2004. A system for visualizing time varying oceanographic 3D data. In *Oceans '04 Mts/leee Techno-Ocean '04, Vols 1- 2, Conference Proceedings, Vols. 1-4*. Kobe, Japan.
- Babak, O., et C. V. Deutsch. 2008. Uncertainty as the overlap of alternate conditional distributions. *Computational Geosciences* 12 (4):503-512.
- Babak, O., et C. V. Deutsch. 2009. Statistical approach to inverse distance interpolation. *Stochastic Environmental Research and Risk Assessment* 23 (5):543-553.
- Basu, A. 1996. Development of a user friendly marine geographic information system with spatial data analysis and error estimation of three dimensional data model in the Hawaiian Exclusive Economic Zone. Paper read at Oceans '96 Mts/leee, Conference Proceedings, Vols 1-3 / Supplementary Proceedings: Coastal Ocean - Prospects for the 21st Century.
- Bathmann, U., M. H. Bundy, M. E. Clarke, T. J. Cowles, K. Daly, H. G. Dam, M. M. Deksheniaks, P. L. Donaghay, D. M. Gibson, D. J. Gifford, B. W. Hansen, D. K. Hartline, E. J. H. Head, E. E. Hofmann, R. R. Hopcroft, R. A. Jahnke, S. H. Jonasdottir, T. Kiorboe, G. S. Kleppel, J. M. Klinck, P. M. Kremer, M. R. Landry, R. F. Lee, P. H. Lenz, L. P. Madin, D. T. Manahan, M. G. Mazzocchi, D. J. McGillicuddy, C. B. Miller, J. R. Nelson, T. R. Osborn, G. A. Paffenhofer, R. E. Pieper, I. Prusova, M. R. Roman, S. Schiel, H. E. Seim, S. L. Smith, J. J. Torres, P. G. Verity, S. G. Wakeham, et K. F. Wishner. 2001. Future marine zooplankton research - a perspective. *Marine Ecology-Progress Series* 222:297-308.
- Behrenfeld, M. J., et E. Boss. 2003. The beam attenuation to chlorophyll ratio: an optical index of phytoplankton physiology in the surface ocean? *Deep-Sea Research Part I-Oceanographic Research Papers* 50 (12):1537-1549.
- Behrenfeld, M. J., et E. Boss. 2006. Beam attenuation and chlorophyll concentration as alternative optical indices of phytoplankton biomass. *Journal of Marine Research* 64 (3):431-451.
- Bergacademie, TU. XXXX. Introduction to 3D Geomodelling with gOcad, lecture materials. Freiberg: Department of Geophysics and Geoinformatics.
- Bond, C.E., A.D. Gibbs, et S. Jones. 2007. What do you think this is? "Conceptual uncertainty" in geoscience interpretation. *GSA Today* 17 (11):4-10.
- Burrough, P.A., et R.A. Mcdonnell. 1998. *Principles of Geographical Information Systems*. Oxford: Oxford University Press.
- Cabello, P., J. L. Cuevas, et E. Ramos. 2007. 3D modelling of grain size distribution in Quaternary deltaic deposits (Llobregat Delta, NE Spain). *Geologica Acta* 5 (3):231-244.

- Carette, V. 2008. Amélioration de la représentation géométrique 2D et 3D des agrégations de poissons en support à l'étude de leur évolution spatio-temporelle. M.Sc diss., Département des sciences géomatiques, Université Laval, Québec.
- Carette, V., M. A. Mostafavi, et R. Devillers. 2008. Towards marine geographic information systems: multidimensional representation of fish aggregations and their spatiotemporal evolutions. In *Oceans 2008, Vols 1-4*. Québec, Canada: MTS/IEEE.
- Carette, V., M. A. Mostafavi, et R. Devillers. 2008. Towards Marine Geographic Information Systems: Multidimensional Representation of Fish aggregations and Their Spatiotemporal Evolutions. In *Oceans 2008, Vols 1-4*.
- Coley, A. R., et P. Claburn. 2005. GIS visualisation and analysis of mobile hydroacoustic fisheries data: a practical example. *Fisheries Management and Ecology* 12 (6):361-367.
- Creel, L. 2003. Ripple effects: population and coastal regions. edited by P. R. Bureau. Washington, USA.
- Cressie, N. 1985. Fitting variogram models by weighted least-squares. *Journal of the International Association for Mathematical Geology* 17 (5):563-586.
- Cressie, N., et D. L. Zimmerman. 1992. On the stability of the geostatistical method. *Mathematical Geology* 24 (1):45-59.
- Cromley, R. G. 1992. *Digital Cartography* Englewood Cliffs, New Jersey: Prentice-Hall.
- Dahle, P., O. Kolbjornsen, et P. Abrahamsen. 2005. When can shape and scale parameters of a 3D variogram be estimated? In *Geostatistics Banff 2004*, edited by O. Leuangthong et V. C. Deutsch. Banff, Canada: Springer.
- Daly, C. 2006. Guidelines for assessing the suitability of spatial climate data sets. *International Journal of Climatology* 26 (6):707-721.
- Daumas, F. 1982. Méthodes de normalisation de données. *Revue de statistique appliqué* 30 (4):23-28.
- DeMers, M. N. 2002. *GIS Modeling in raster*. John Wiley and Sons, INC.
- Desassis, N., et D. Renard. 2012. Automatic Variogram Modeling by Iterative Least Squares: Univariate and Multivariate Cases. *Journal of Mathematical Geology* 45 (4):453-470.
- Deutsch, C. V. 1994. Kriging with strings of data. *Mathematical Geology* 26 (5):623-638.
- Deutsch, C.V. 2002. *Geostatistical reservoir modeling*. Edited by A. G. Journel: Oxford University Press.
- Falivene, O., L. Cabrera, J. A. Munoz, P. Arbues, O. Fernandez, et A. Saez. 2007. Statistical grid-based facies reconstruction and modelling for sedimentary bodies. Alluvial-palustrine and turbiditic examples. *Geologica Acta* 5 (3):199-230.
- Falivene, O., L. Cabrera, et A. Saez. 2007. Optimum and robust 3D facies interpolation strategies in a heterogeneous coal zone (Tertiary As Pontes basin, NW Spain). *International Journal of Coal Geology* 71 (2-3):185-208.
- Falivene, O., L. Cabrera, R. Tolosana-Delgado, et A. Saez. 2010. Interpolation algorithm ranking using cross-validation and the role of smoothing effect. A coal zone example. *Computers & Geosciences* 36 (4):512-519.
- Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, et P. Burlando. 2007. Testing alternative ground water models using cross-validation and other methods. *Ground Water* 45 (5):627-641.
- Forest, A, M. Babin, L. Stemmann, M. Picheral, M. Sampei, L. Fortier, Y. Gratton, S. Bélanger, E. Devred, J. Sahlin, D. Doxaran, F. Joux, E. Ortega-Retuerta, J. Martin, W. H. Jeffrey, B. Gasser, et J. Carlos Miquel. 2013. Ecosystem function and particle flux dynamics across the Mackenzie Shelf (Beaufort Sea, Arctic Ocean): an integrative analysis of spatial variability and biophysical forcings *Biogeosciences* 10 (5):2833-2866.

- Forest, A., M. Sampei, H. Hattori, R. Makabe, H. Sasaki, M. Fukuchi, P. Wassmann, et L. Fortier. 2007. Particulate organic carbon fluxes on the slope of the Mackenzie Shelf (Beaufort Sea): Physical and biological forcing of shelf-basin exchanges. *Journal of Marine Systems* 68 (1-2):39-54.
- Forest, A., L. Stemann, M. Picheral, L. Burdorf, D. Robert, L. Fortier, et M. Babin. 2012. Size distribution of particles and zooplankton across the shelf-basin system in southeast Beaufort Sea: combined results from an Underwater Vision Profiler and vertical net tows. *Biogeosciences* 9 (4):1301-1320.
- Francois, R., M. Amini, E. Asher, I. Beveridge, K. Brown, J. Carpenter, J. Cullen, B. De Baere, C. Guignard, M. Hale, A. Hamilton, M. Hernandez, A. Kobryn, M. Maldonado, J. McAllister, K. Orians, C. Payne, R. Ramirez, R. Rivkin, D. Semeniuk, M. Soon, E. Sternberg, N. Sutherland, R. Taylor, et D. Varela. 2009. GEOTRACES: Multi-tracer investigation of the effect of climate change on nutrient and carbon cycles in the Arctic Ocean, ArcticNet 0903-LEG 3a, August 27-September 12, 2009, Paulatuk-Paulatuk, CCGS Amundsen.
- Game, E. T., H. S. Grantham, A. J. Hobday, R. L. Pressey, A. T. Lombard, L. E. Beckley, K. Gjerde, R. Bustamante, H. P. Possingham, et A. J. Richardson. 2009. Pelagic protected areas: the missing dimension in ocean conservation. *Trends in Ecology & Evolution* 24 (7):360-369.
- Gold, C., et A.R. Condal. 1995. A spatial data structure Integrating GIS and Simulation in a Marine Environment. *Marine Geodesy* 18:pp. 213-228.
- Gold, C. M., et A. R. Condal. 1995. A spatial data structure integrating GIS and simulation in a marine environment. *Marine Geodesy* 18 (3):213-228.
- Goodchild, M. F. 1992. Geographical Data Modeling. *Computers & Geosciences* 18 (4):401-408.
- Goovaerts, P. 1997. *Geostatistics for natural resources evaluation, Applied Geostatistics Series*. New York: Oxford University Press.
- Gotway, C. A., R. B. Ferguson, G. W. Hergert, et T. A. Peterson. 1996. Comparison of kriging and inverse-distance methods for mapping soil parameters. *Soil Science Society of America Journal* 60 (4):1237-1247.
- Gringarten, E., et C. V. Deutsch. 2001. Variogram interpretation and modeling. *Mathematical Geology* 33 (4):507-534.
- Harris, P. T., A. D. Heap, T. Whiteway, et A. Post. 2008. Application of biophysical information to support Australia's representative marine protected area program. *Ocean & Coastal Management* 51 (10):701-711.
- Hashemi Beni, L., M. A Mostafavi, J Pouliot, et M Gavrilova. 2011. Toward 3D spatial dynamic field simulation within GIS using kinetic Voronoi diagram and Delaunay tetrahedralization. *International Journal of Geographical Information Science* 25 (1):25-50.
- Hashemi Beni, L. 2009. Development of a 3D kinetic data structure adapted for a 3D spatial dynamic field simulation, Faculté des études supérieures de l'Université Laval, Laval University, Québec
- Hashemi Beni, L. 2009. Development of a 3D kinetic data structure adapted for a 3D spatial dynamic field simulation, Dep. des sciences géomatiques, Université Laval, Québec.
- He, Y., D. Chen, B. G. Li, Y. F. Huang, K. L. Hu, Y. Li, et I. R. Willett. 2009. Sequential indicator simulation and indicator kriging estimation of 3-dimensional soil textures. *Australian Journal of Soil Research* 47 (6):622-631.
- He, Y., K. L. Hu, D. L. Chen, H. C. Suter, Y. Li, B. G. Li, X. Y. Yuan, et Y. F. Huang. 2010. Three dimensional spatial distribution modeling of soil texture under agricultural systems using a sequence indicator simulation algorithm. *Computers and Electronics in Agriculture* 71:S24-S31.

- He, Y., K. L. Hu, Y. F. Huang, B. G. Li, et D. L. Chen. 2010. Analysis of the anisotropic spatial variability and three-dimensional computer simulation of agricultural soil bulk density in an alluvial plain of north China. *Mathematical and Computer Modelling* 51 (11-12):1351-1356.
- Head, M. E. M., P. Luong, J. H. Costolo, K. Countryman, et C. Szczechowski. 1997. Applications of 3-D visualizations of oceanographic data bases. In *Oceans '97 MTS/IEEE*. Halifax, Canada: IEEE.
- Isaaks, E. H., et G. Srivastava. 1992. *An introduction to applied geostatistics*. New York, USA: Oxford University Press.
- Isaaks, E. H., et R. M. Srivastava. 1988. Spatial continuity measures for probabilistic and deterministic geostatistics. *Mathematical Geology* 20 (4):313-341.
- Jakobsson, M., L. Macnab, R. Mayer, L. Anderson, M. Edwards, J. Hatzky, H. W. Schenke, et P. Johnson. 2008. An improved bathymetric portrayal of the Arctic Ocean: Implications for ocean modeling and geological, geophysical and oceanographic analyses. *Geophysical Research Letters* 35 (7):1-5.
- Jensen, O. P., M. C. Christman, et T. J. Miller. 2006. Landscape-based geostatistics: a case study of the distribution of blue crab in Chesapeake Bay. *Environmetrics* 17 (6):605-621.
- Jones, R. R., K. J. W. McCaffrey, P. Clegg, R. W. Wilson, N. S. Holliman, R. E. Holdsworth, J. Imber, et S. Waggott. 2009. Integration of regional to outcrop digital data: 3D visualisation of multi-scale geological models. *Computers & Geosciences* 35 (1):4-18.
- Kaufman, A. E. 1996. Volume visualization. *Acm Computing Surveys* 28 (1):165-167.
- Kelkar, M., et G. Perez. 2002. *Applied geostatistics for reservoir characterization*. Edited by A. Chopra. Richardson, Texas: Society of Petroleum Engineers.
- Kerry, R., et M. A. Oliver. 2007. Determining the effect of asymmetric data on the variogram I. Underlying asymmetry. *Computers & Geosciences* 33 (10):1212-1232.
- Kracker, L. M. 1999. The geography of fish: The use of remote sensing and spatial analysis tools in fisheries research. *Professional Geographer* 51 (3):440-450.
- Kravchenko, A., et D. G. Bullock. 1999. A comparative study of interpolation methods for mapping soil properties. *Agronomy Journal* 91 (3):393-400.
- Lahiri, S. N., Y. D. Lee, et N. Cressie. 2002. On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *Journal of Statistical Planning and Inference* 103 (1-2):65-85.
- Lam, N. S. N. 1983. Spatial Interpolation Methods- A review. *American Cartographer* 10 (2):129-149.
- Lansard, B., A. Mucci, L. A. Miller, R. W. Macdonald, et Y. Gratton. 2012. Seasonal variability of water mass distribution in the southeastern Beaufort Sea determined by total alkalinity and $\delta^{18}\text{O}$. *Journal of Geophysical Research* 117:1-19.
- Ledoux, H. 2006. Modelling three-dimensional geoscientific fields with the Voronoi diagram and its dual, School of Computing, University of Glamorgan, Pontypridd, Wales, UK.
- Ledoux, H., et C. Gold. 2006. La modélisation de données océanographiques à l'aide du diagramme Voronoi. *Géomatique* 16/2006:51-70.
- Ledoux, H., et C. M. Gold. 2006. A Voronoi-based map algebra. In *Progress in Spatial Data Handling-12th International Symposium on Spatial Data Handling*, edited by A. Reidl, W. Kainz et G. Elmes: Springer.
- Ledoux, H., et C. M. Gold. 2008. Modelling three-dimensional geoscientific fields with the Voronoi diagram and its dual. *International Journal of Geographical Information Science* 22 (5):547-574.
- Li, C. K., L. Chen, Y. Wang, et S. N. Zheng. 2008. *On effects of weights in spatial interpolation*. Edited by J. X. Zhang et M. F. Goodchild. Vol. 1, *Proceedings of the 8th International Symposium on Spatial Accuracy*

- Assessment in Natural Resources and Environmental Sciences*. Shanghai, People's Republic of China: World Acad Press.
- Li, J., et A. D. Heap. 2008. A review of Spatial Interpolation Methods for Environmental Scientists. Australian Government, Geoscience Australia, Record: 2008/23.
- Li, J., et A. D. Heap. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* 6 (3-4):228-241.
- Li, J., A. D. Heap, A. Potter, et J. J. Daniell. 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environmental Modelling & Software* 26 (12):1647-1659.
- Li, R. 1994. Data Structure and application issue in 3-D Geographic Information Systems. *Geomatica* 48 (3):209-224.
- Li, W., W. L. Liu, Y. Zhang, W. Yin, et Y. Liu. 2010. On the comparison of spatial interpolation methods of marine temperature and salinity based on Arcgis software: a case study of Tianjin coastal waters in the Bohai bay. In *Sixth International Symposium on Digital Earth: Models, Algorithms, and Virtual Reality*, edited by H. Guo et C. Wang. Beijing, People's Republic of China: SPIE-Int Soc Optical Engineering.
- MacCormack, K. E., J. J. Brodeur, et C. H. Eyles. 2013. Evaluating the impact of data quantity, distribution and algorithm selection on the accuracy of 3D subsurface models using synthetic grid models of varying complexity. *Journal of Geographical Systems* 15 (1):71-88.
- MacCormack, K. E., et C. H. Eyles. 2012. Assessing the impact of program selection on the accuracy of 3D geologic models. *Geosphere* 8 (2):534-543.
- Macdonald, R. W., S. M. Solomon, R. E. Cranston, H. E. Welch, M. B. Yunker, et C. Gobeil. 1998. A sediment and organic carbon budget for the Canadian Beaufort shelf. *Marine Geology* 144 (4):255-273.
- Malina. 2013. *Description of the Beaufort Sea study area* 2013 [cited april 28, 2013 2013]. Available from http://malina.obs-mfr.fr/cruise_description.html?tab=0#phyto.
- Martin, A. P., K. J. Richards, A. Bracco, et A. Provenzale. 2002. Patchy productivity in the open ocean. *Global Biogeochemical Cycles* 16 (2).
- Martin, J., J. E. Tremblay, J. Gagnon, G. Tremblay, A. Lapoussiere, C. Jose, M. Poulin, M. Gosselin, Y. Gratton, et C. Michel. 2010. Prevalence, structure and properties of subsurface chlorophyll maxima in Canadian Arctic waters. *Marine Ecology-Progress Series* 412:69-84.
- Matsuoka, A., A. Bricaud, R. Benner, J. Para, R. Sempéré, L. Prieur, S. Bélanger, et M. Babin. 2012. Tracing the transport of colored dissolved organic matter in water masses of the Southern Beaufort Sea: relationship with hydrographic characteristics. *Biogeosciences* 9:925-940.
- Mesick, S. M., T. C. Vance, C. J. Beegle-Krauss, et D. Steube. 2009. HabitatSpace: Multidimensional Characterization of Pelagic Essential Fish Habitat. In *Oceans 2009, Vols 1-3*. New York: IEEE.
- Minster, J.-F. 1997. *La Machine-Océan*. Paris: Flammarion.
- Miranda, H., et M. S. de Miranda. 2011. Combining Robustness with Efficiency in the Estimation of the Variogram. *Mathematical Geosciences* 43 (3):363-377.
- Mostafavi, M. A. 2002. Development of a global dynamic data structure, Département des sciences géomatiques, Université Laval, Québec.
- Mostafavi, M. A., et C. Gold. 2004. A global kinetic spatial data structure for a marine simulation. *International Journal of Geographical Information Science* 18 (3):211-227.
- Mucha, J., et M. Wasilewska-Blaszczyk. 2012. Variability anisotropy of mineral deposits parameters and its impact on resources estimation - a geostatistical approach. *Gospodarka Surowcami Mineralnymi-Mineral Resources Management* 28 (4):113-135.

- Mueller, T. G., N. B. Pusuluri, K. K. Mathias, P. L. Cornelius, R. I. Barnhisel, et S. A. Shearer. 2004. Map quality for ordinary kriging and inverse distance weighted interpolation. *Soil Science Society of America Journal* 68 (6):2042-2047.
- Muller, W. G., et D. L. Zimmerman. 1999. Optimal designs for variogram estimation. *Environmetrics* 10 (1):23-37.
- Murphy, R. R., F. C. Curriero, et W. P. Ball. 2010. Comparison of Spatial Interpolation Methods for Water Quality Evaluation in the Chesapeake Bay. *Journal of Environmental Engineering-Asce* 136 (2):160-171.
- Ng, S. M. Y., O. W. H. Wai, Z. H. Xu, Y. S. Li, H. L. Chen, et Z. L. Li. 2007. Development of a GIS for managing dynamic, 3D coastal information of Pearl River Estuary. *Journal of Hydroinformatics* 9 (3):215-232.
- Olea, R. A. 2007. Declustering of clustered preferential sampling for histogram and semivariogram inference. *Mathematical Geology* 39 (5):453-467.
- OQLF. 2013. *Office québécois de la langue française, 1997* Gouvernement du Québec 2013 [cited 7 mai 2013 2013]. Available from http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2076213.
- Park, J. J., K. I. Shin, J. H. Lee, S. E. Lee, W. K. Lee, et K. Cho. 2012. Detecting and cleaning outliers for robust estimation of variogram models in insect count data. *Ecological Research* 27 (1):1-13.
- Pauly, D., V. Christensen, S. Guenette, T. J. Pitcher, U. R. Sumaila, C. J. Walters, R. Watson, et D. Zeller. 2002. Towards sustainability in world fisheries. *Nature* 418 (6898):689-695.
- Peuquet, D. J. 1988. Representations of Geographic space- toward a conceptual synthesis. *Annals of the Association of American Geographers* 78 (3):375-394.
- Picheral, M., L. Guidi, L. Stemann, D. M. Karl, G. Iddaoud, et G. Gorsky. 2010. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. *Limnology and Oceanography-Methods* 8:462-473.
- Pouliot, J., K. Bedard, D. Kirkwood, et B. Lachance. 2008. Reasoning about geological space: Coupling 3D GeoModels and topological queries as an aid to spatial data selection. *Computers & Geosciences* 34 (5):529-541.
- Pouliot, J., B. Lachance, et D. Kirkwood. 2006. L'importance de la modélisation géométrique 3D lors de l'élaboration d'un SIG 3D: Exemple du développement d'une structure topologique pour une application géologique. *Revue Internationale de Géomatique* 16 (1):29-49.
- Raiber, M., P. A. White, C. J. Daughney, C. Tschirter, P. Davidson, et S. E. Bainbridge. 2012. Three-dimensional geological modelling and multivariate statistical analysis of water chemistry data to analyse and visualise aquifer structure and groundwater composition in the Wairau Plain, Marlborough District, New Zealand. *Journal of Hydrology* 436:13-34.
- Reed, P. M., T. R. Ellsworth, et B. S. Minsker. 2004. Spatial interpolation methods for nonstationary plume data. *Ground Water* 42 (2):190-202.
- Rivest, M., D. Marcotte, et P. Pasquier. 2012. Sparse data integration for the interpolation of concentration measurements using kriging in natural coordinates. *Journal of Hydrology* 416:72-82.
- Robins, N. S., H. K. Rutter, S. Dumbleton, et D. W. Peach. 2005. The role of 3D visualisation as an analytical tool preparatory to numerical modelling. *Journal of Hydrology* 301 (1-4):287-295.
- Sahlin, J., M. A. Mostafavi, A. Forest, M. Babin, et B. Lansard. 2012. 3D geospatial modelling and visualization for marine environment: Study of the marine pelagic ecosystem of the south-eastern Beaufort Sea, Canadian Arctic. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XXXVIII-4/C26, 21-24.
- Sampei, M., H. Sasaki, R. Makabe, A. Forest, H. Hattori, J. E. Tremblay, Y. Gratton, M. Fukuchi, et L. Fortier. 2011. Production and retention of biogenic matter in the southeast Beaufort Sea during 2003-2004:

- insights from annual vertical particle fluxes of organic carbon and biogenic silica. *Polar Biology* 34 (4):501-511.
- Schloeder, C. A., N. E. Zimmerman, et M. J. Jacobs. 2001. Comparison of methods for interpolating soil properties using limited data. *Soil Science Society of America Journal* 65 (2):470-479.
- Shyue, S. W., et P. Y. Tsai. 1996. A study on the dimensional aspect of the marine geographic information systems. In *Prospects for the 21st-Century, OCEANS 96 MTS/IEEE*. Lauredale, USA: IEEE.
- Tobler, W.R. 1970. A computer movie simulating urban growth in the Detroit Region. *Economic Geography* 46:234-240.
- Tomlin, C. D. 1994. MAP ALGEBRA - ONE PERSPECTIVE. *Landscape and Urban Planning* 30 (1-2):3-12.
- Tremblay, J. E., S. Belanger, D. G. Barber, M. Asplin, J. Martin, G. Darnis, L. Fortier, Y. Gratton, H. Link, P. Archambault, A. Sallon, C. Michel, W. J. Williams, B. Philippe, et M. Gosselin. 2011. Climate forcing multiplies biological productivity in the coastal Arctic Ocean. *Geophysical Research Letters* 38 (L18604):1-5.
- Tse, R. O. C., et C. Gold. 2004. TIN meets CAD - extending the TIN concept in GIS. *Future Generation Computer Systems* 20 (7):1171-1184.
- Valdés, L., L. Fonseca, et K. Tedesco. 2010. Looking into the future of ocean sciences: An IOC perspective. *Oceanography* 23 (3):160-175.
- van der Putt, T. 2009. Using the discrete 3D Voronoi diagram for the modelling of 3D continuous information in geosciences, Geographical Information Management and Applications, Utrecht University, Utrecht.
- van Ruth, P. D., G. G. Ganf, et T. M. Ward. 2010. Hot-spots of primary productivity: An Alternative interpretation to Conventional upwelling models. *Estuarine Coastal and Shelf Science* 90 (3):142-158.
- Vann, J., S. Jackson, et O. Bertoli. 2003. Quantitative kriging neighbourhood analysis for the mining geologist- A description of the method with worked case examples. In *5th International Mining Geology Conference*, edited by S. Dominy. Bendigo, Australia: Australian Inst Mining & Metallurgy.
- Weber, D. D., et E. J. Englund. 1994. Evaluation and comparison of spatial interpolators II. *Mathematical Geology* 26 (5):589-603.
- Weber, D., et E. Englund. 1992. Evaluation and comparison of spatial interpolators I. *Mathematical Geology* 24 (4):381-391.
- Webster, R. , et M. Oliver. 2007. *Geostatistics for Environmental Scientists*. 2nd edition ed. Chichester: John Wiley & Sons, Ltd.
- Williams, W. J., et E. C. Carmack. 2008. Combined effect of wind-forcing and isobath divergence on upwelling at Cape Bathurst, Beaufort Sea. *Journal of Marine Research* 66 (5):645-663.
- Willmott, C. J., et K. Matsuura. 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research* 30 (1):79-82.
- Wright, D. J., et M. F. Goodchild. 1997. Data from the deep: Implications for the GIS community. *International Journal of Geographical Information Science* 11 (6):523-528.
- Wright, D.J, et D.J. Bartlett. 2001. *Marine and coastal geographical information systems*: Taylor and Francis.
- Wycisk, P., T. Hubert, W. Gossel, et C. Neumann. 2009. High-resolution 3D spatial modelling of complex geological structures for an environmental risk assessment of abundant mining and industrial megasites. *Computers & Geosciences* 35 (1):165-182.
- Zheng, Kai, Xiaofang Zhou, Pui Cheong Fung, et Kexin Xie. 2012. Spatial query processing for fuzzy objects. *Vldb Journal* 21 (5):729-751.
- Zhou, Feng, Huai-Cheng Guo, Yuh-Shan Ho, et Chao-Zhong Wu. 2007. Scientometric analysis of geostatistics using multivariate methods. *Scientometrics* 73 (3):265-279.

Zimmerman, D., C. Pavik, A. Ruggles, et M. P. Armstrong. 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology* 31 (4):375-390.

Annexe A. Variogrammes 3D.

Cette annexe présente les variogrammes 3D anisotropiques pour les 5 variables évaluées. Variogrammes sont modélisés dans quatre intervalles d'azimut (azimut central : 0°, 45°, 90°, 135°) et trois intervalles d'angles de rotation verticale (0°, 45°, 85°). Modèle du variogramme = sphérique. Les variables ont été transformées en une distribution gaussienne normalisée L'axe Y correspond au γ . L'axe X représente la fraction de l'espace de modélisation (*grid*) utilisée pour la modélisation du variogramme. Variogramme pour c_p est présenté dans le texte (section 2.2).

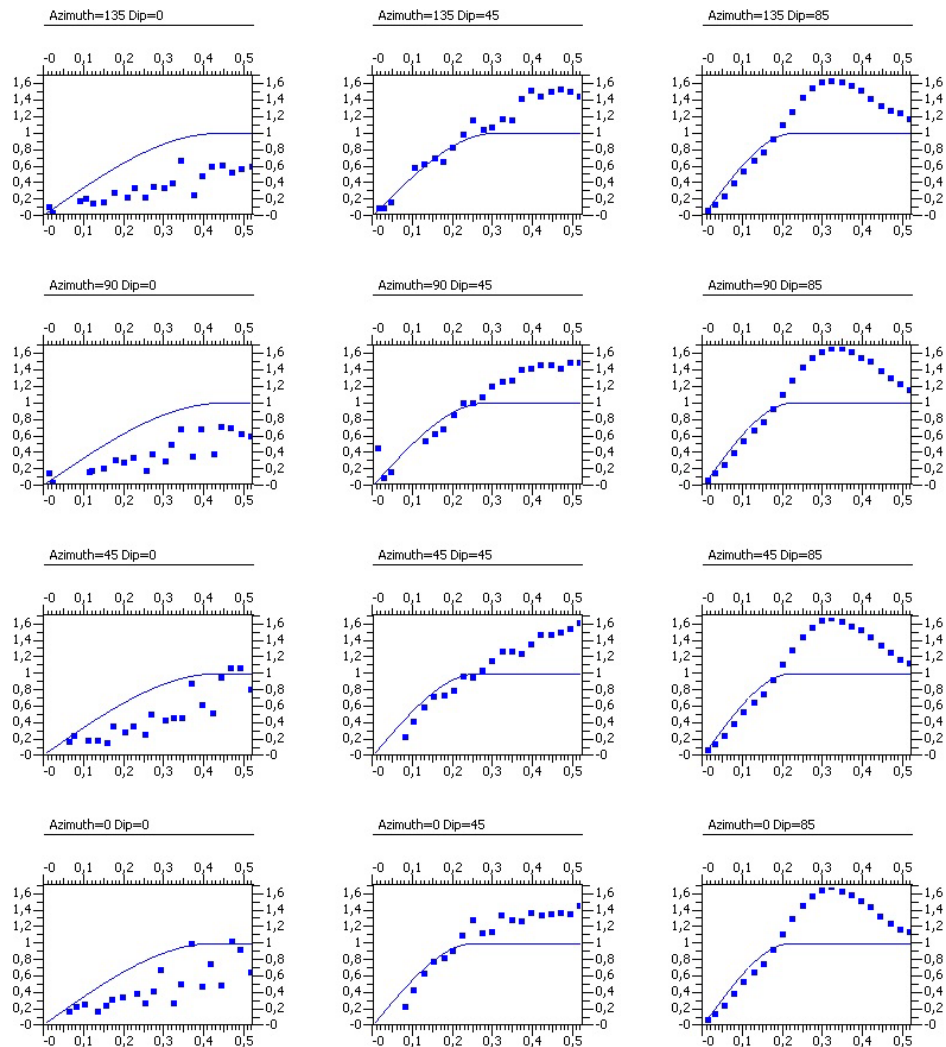


Figure 28. Variogramme pour température (sous-ensemble).

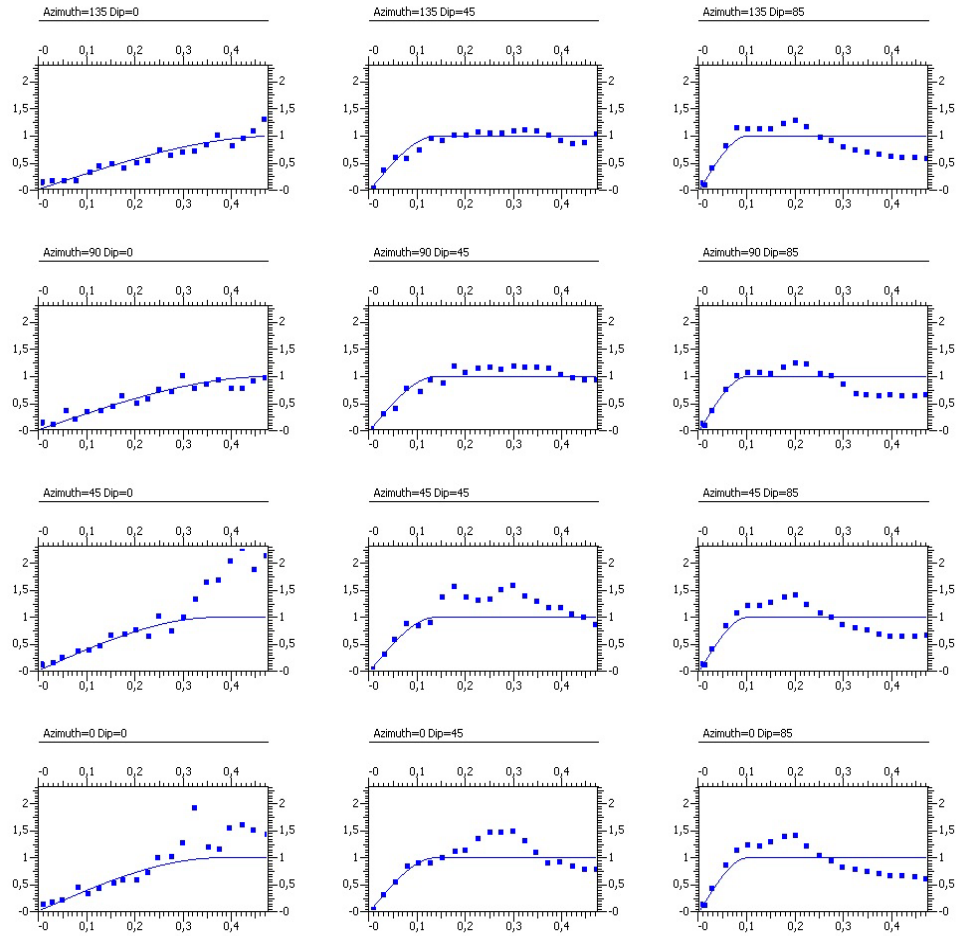


Figure 29. Variogramme pour température (jeu de données complet).

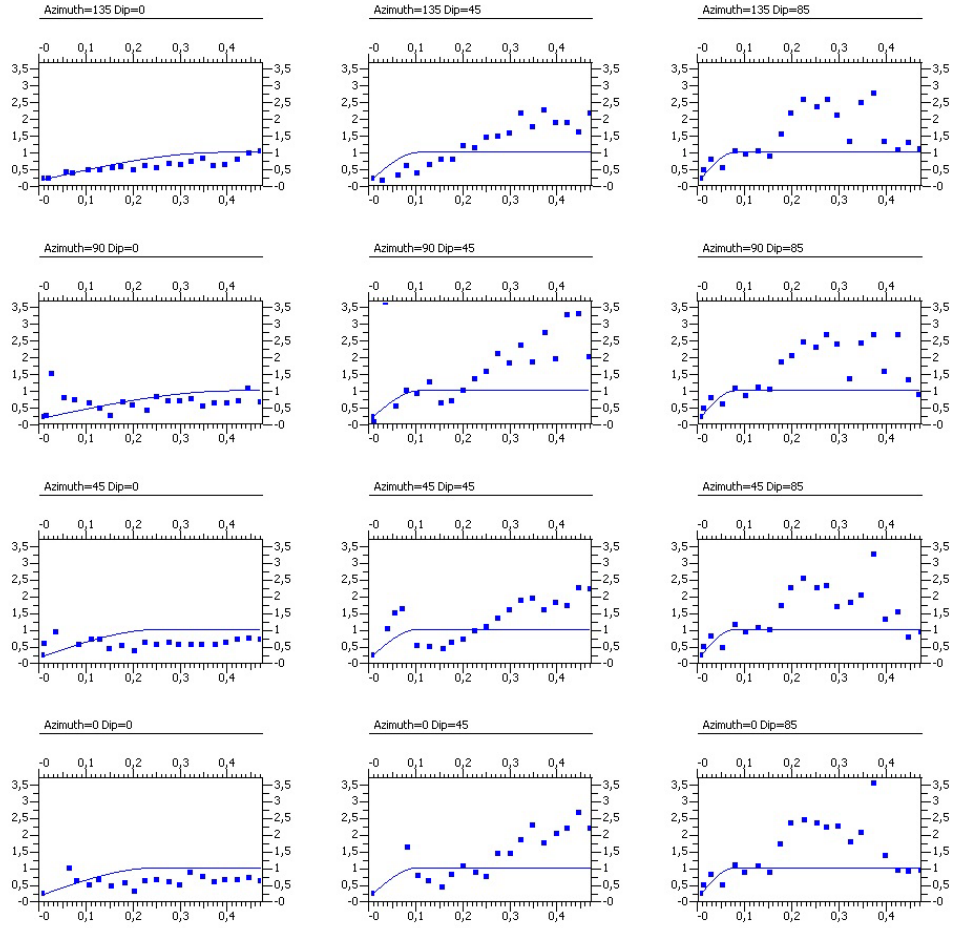


Figure 30. Variogramme pour UHW (*grid* parallel).

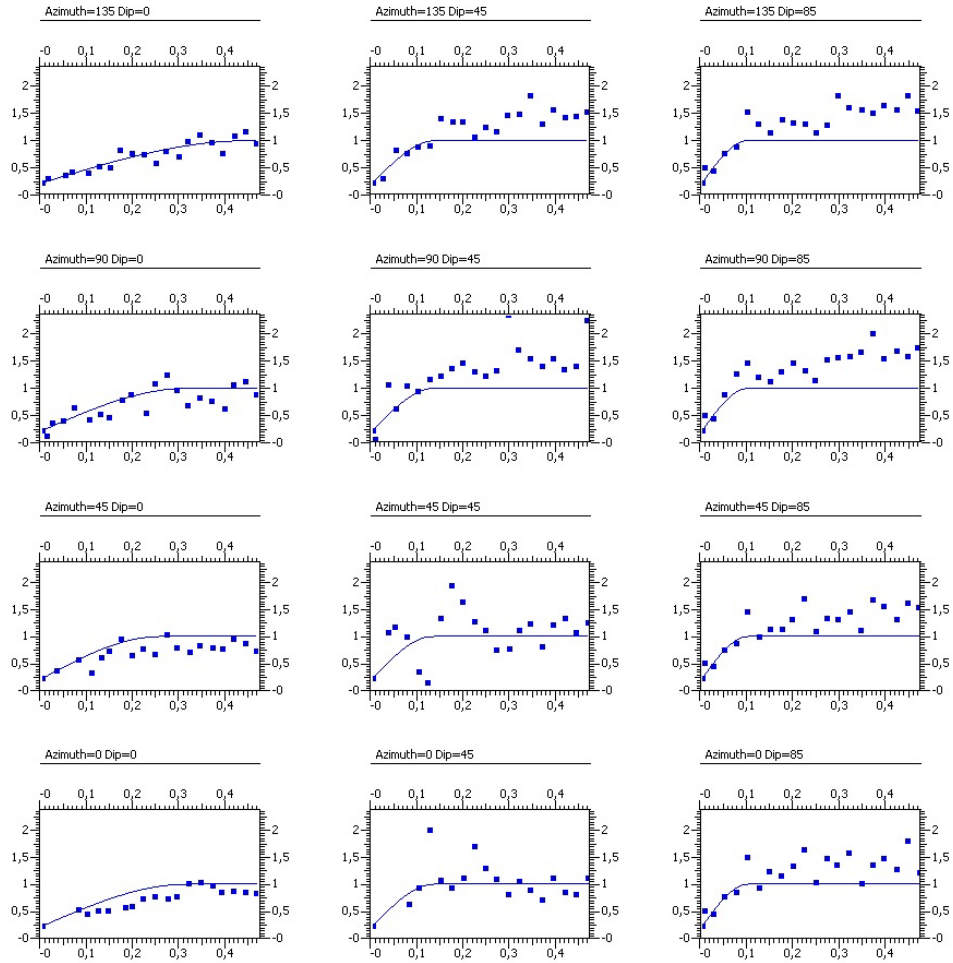


Figure 31. Variogramme pour UHW (*grid proportionnel*).

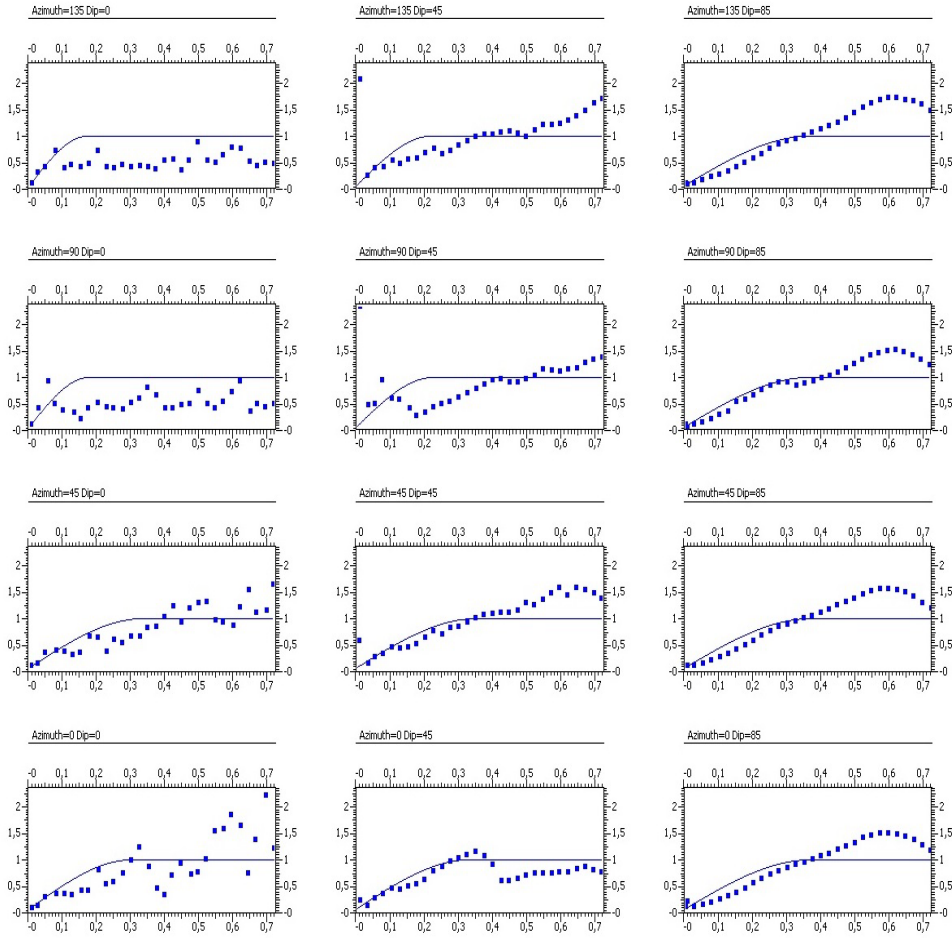


Figure 32. Variogramme pour chl a.

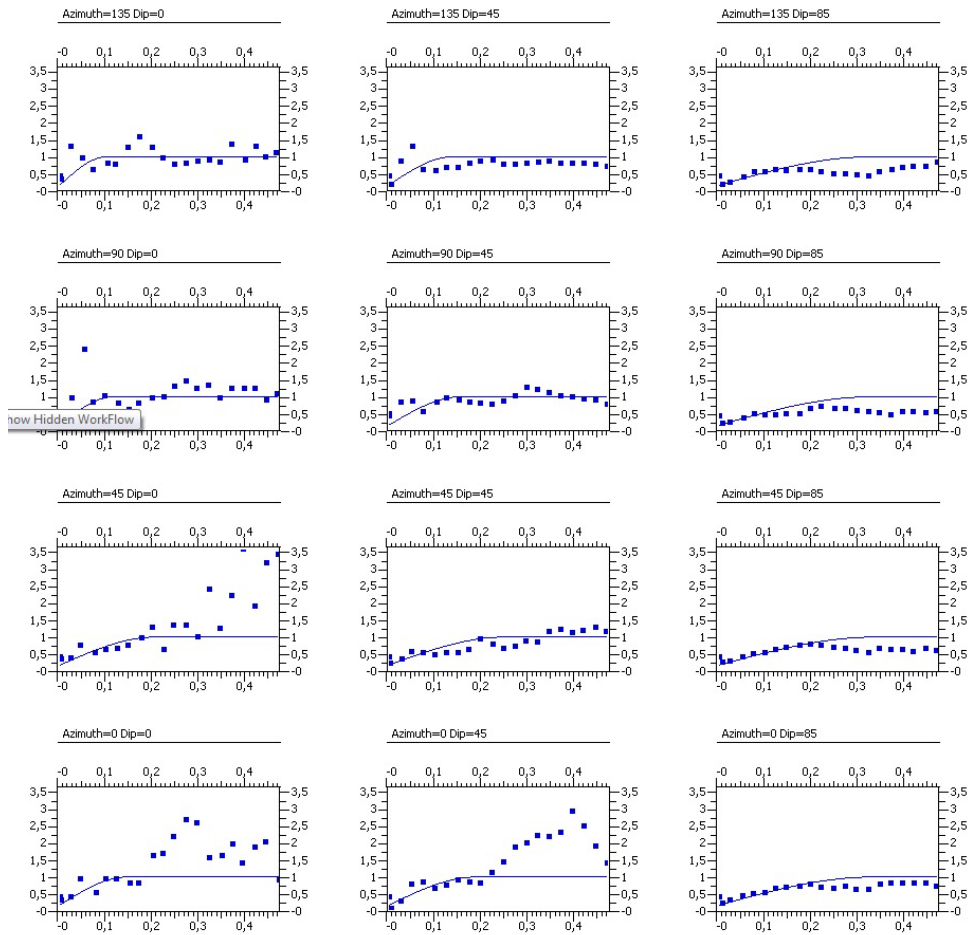


Figure 33. Variogramme pour flux vertical de COP.

Annexe B. Résultats de la validation croisée (CV) des stratégies d'interpolation.

Tableau 6. Performance des méthodes d'interpolation 3D sur la base de métriques d'erreur MSE et MAE. n représente le voisinage de l'interpolation, les valeurs de MSE et MAE correspondent au métrique d'erreur moyenne d'une interpolation par rapport aux valeurs mesurées (voir texte section 4.5.3.).

		Résultats CV- MSE						Résultats CV- MAE					
Stratégie d'interpolation	n	T_{subset}	T.	UHW	chl a	C_p	COP _{flux}	T_{subset}	T.	UHW	chl a	C_p	COP _{flux}
Contrôle (NN)	1	0.274	1.752	218	0.754	0.053	52862	0.265	0.671	7.232	0.281	0.087	110.10
OK	8			111						5.686			
	16	0.072	1.363	109	0.616	0.054	29522	0.175	0.614	5.837	0.267	0.091	89.65
	32	0.067	1.359	106	0.610	0.055	28959	0.173	0.622	5.824	0.269	0.092	88.40
	48	0.066	1.358		0.601	0.054	29023	0.173	0.622		0.270	0.094	88.83
	96	0.067	1.367		0.608	0.054	28708	0.173	0.628		0.273	0.094	88.48
OK - Sous-estimation des portées horizontales	8			114						5.568			
	16	0.083	1.357	114	0.627	0.048		0.197	0.605	5.642	0.274	0.090	
	32	0.095	1.350	114	0.642	0.049		0.217	0.605	5.566	0.279	0.091	
	48	0.101	1.346		0.647	0.049		0.223	0.605		0.283	0.092	
	96	0.103	1.361		0.650	0.049		0.224	0.612		0.286	0.091	
OK - Exagération des portées horizontales	8			115						5.595			
	16	0.064	1.336	110	0.606	0.051	31368	0.163	0.602	5.647	0.259	0.089	86.87
	32	0.055	1.306	106	0.600	0.049	31152	0.152	0.586	5.528	0.262	0.089	89.43
	48	0.055	1.331		0.593	0.050	31393	0.150	0.602		0.263	0.090	89.96
	96	0.052	1.327		0.595	0.050	32078	0.148	0.603		0.265	0.090	92.58
IDW _{xyz} Power 1	8		2.165	270	0.675	0.058	28130		0.779	11.772	0.266	0.090	82.45
	16	0.151	2.539	326	0.698	0.065	29421	0.249	0.833	13.914	0.272	0.095	83.62
	32			460						17.681			
	48	0.223	3.186		0.701	0.063	30600	0.367	0.927		0.272	0.089	83.12
	96	0.258	3.392		0.732	0.063	32646	0.417	0.975		0.272	0.087	83.27
IDW _{xyz} Power 2	8		2.001	202	0.668	0.056	28806		0.825	9.673	0.263	0.088	83.31
	16	0.137	2.214	222	0.686	0.061	28563	0.231	0.912	10.734	0.269	0.092	80.97
	32			286						13.034			
	48	0.181	2.684		0.685	0.060	28115	0.318	1.036		0.268	0.088	77.74
	96	0.202	2.918		0.705	0.060	28460	0.350	1.090		0.267	0.087	76.34
IDW _{xyz} Power 5	8		1.767	169	0.662	0.053	31453		0.706	7.628	0.261	0.087	87.66
	16	0.126	1.779	161	0.668	0.054	30618	0.210	0.710	7.647	0.265	0.089	85.62
	32			167						8.107			
	48	0.139	1.806		0.664	0.054	29873	0.252	0.720		0.264	0.087	83.81
	96	0.142	1.823		0.666	0.054	29463	0.260	0.727		0.263	0.087	82.99
IDW _{uw} Power 1	8		2.224	116	0.651	0.057	27990		0.797	8.413	0.261	0.088	81.09
	16		2.503	118	0.695	0.063	29043		0.865	10.815	0.277	0.092	83.22
	32			116						14.860			
	48		3.207		0.703	0.063	30582		0.990		0.273	0.089	84.88
	96		3.585		0.729	0.063	32550		1.059		0.270	0.088	85.50
IDW _{uw} Power 2	8		2.019	121	0.651	0.056	28438		0.758	7.678	0.260	0.087	81.73
	16		2.204	122	0.682	0.059	28335		0.812	9.145	0.272	0.090	80.33
	32			120						11.640			
	48		2.746		0.687	0.060	28067		0.913		0.269	0.088	78.41
	96		3.088		0.704	0.061	28446		0.974		0.266	0.087	77.71
IDW _{uw} Power 5	8		1.756	144	0.656	0.053	30982		0.696	7.093	0.259	0.087	86.28
	16		1.786	143	0.666	0.059	30417		0.712	7.368	0.266	0.090	85.12
	32			142						7.869			
	48		1.831		0.665	0.054	29609		0.728		0.264	0.087	83.45
	96		1.866		0.666	0.054	29332		0.743		0.263	0.087	82.94

Annexe C. Évaluation qualitative des stratégies d'interpolation 3D.

Cette annexe présente les représentations spatiales provenant de chaque stratégie d'interpolation. Une représentation par méthode d'interpolation (OK, IDW_{xyz}, IDW_{uvw}, et NN) est présentée, ce qui correspond dans chaque cas à la stratégie avec les erreurs de CV les plus faibles (donc, la stratégie avec le paramétrage optimal en ce qui concerne les métriques d'erreurs MAE et MSE). Les représentations spatiales de la température ont été présentées auparavant (section 4.6.3, Figure 22). Une visualisation qui optimise la comparaison qualitative a été utilisée pour chaque variable. Pour c_p et chl a , la différence entre les méthodes d'interpolation s'observe bien avec une visualisation des volumes (une grande différence des valeurs à la surface de l'eau). Pour UHW et le flux vertical de COP, la variabilité est mieux illustrée avec des sections verticales. n représente le voisinage de l'interpolation (nombre de points utilisés dans les interpolations) et p correspond au paramètre du puissance dans la méthode IDW. Pour OK, une stratégie avec une exagération de la portée horizontale est indiquée avec OK, « exagg. hor. rang ».

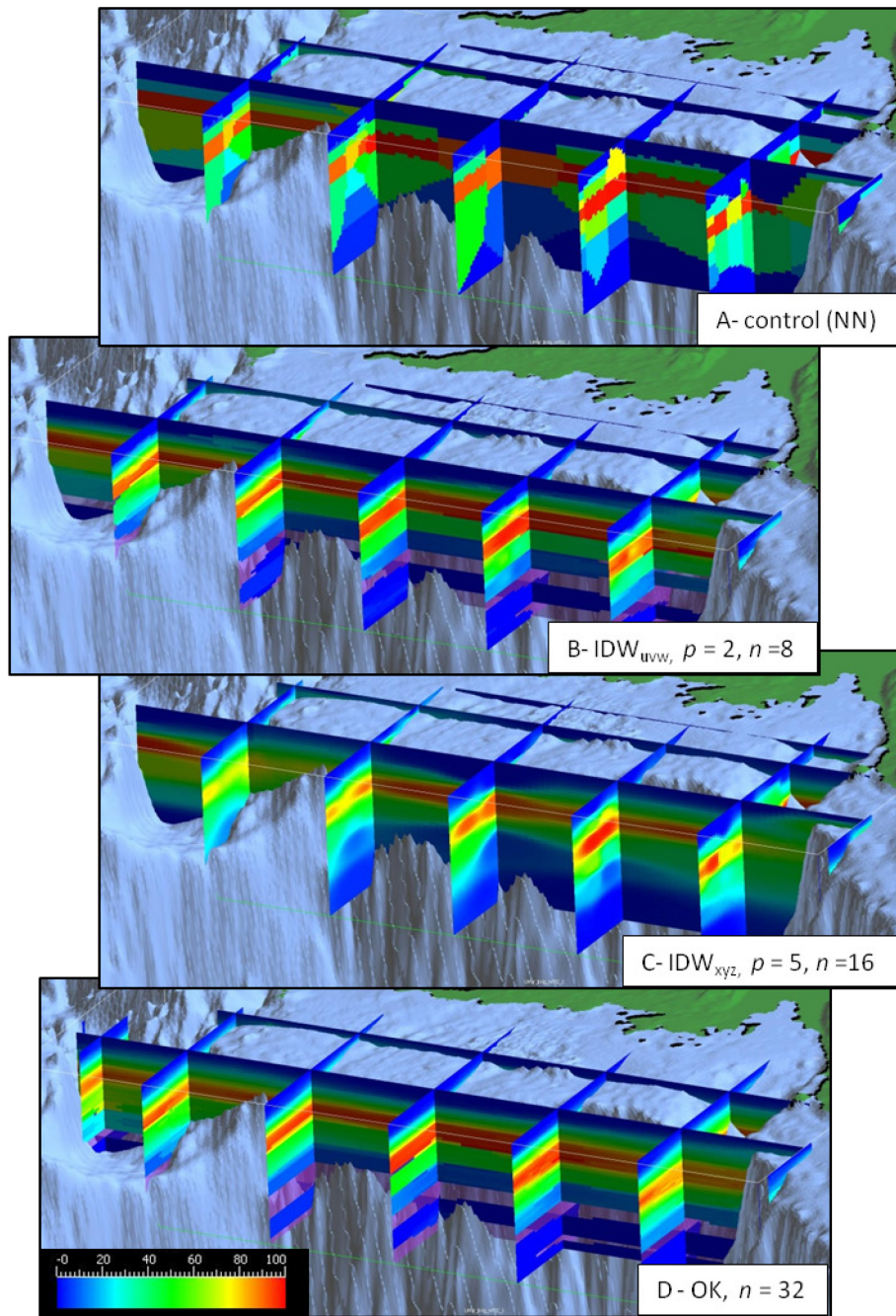


Figure 34. Représentations spatiales 3D de la distribution de UHW (%). Les 500 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200

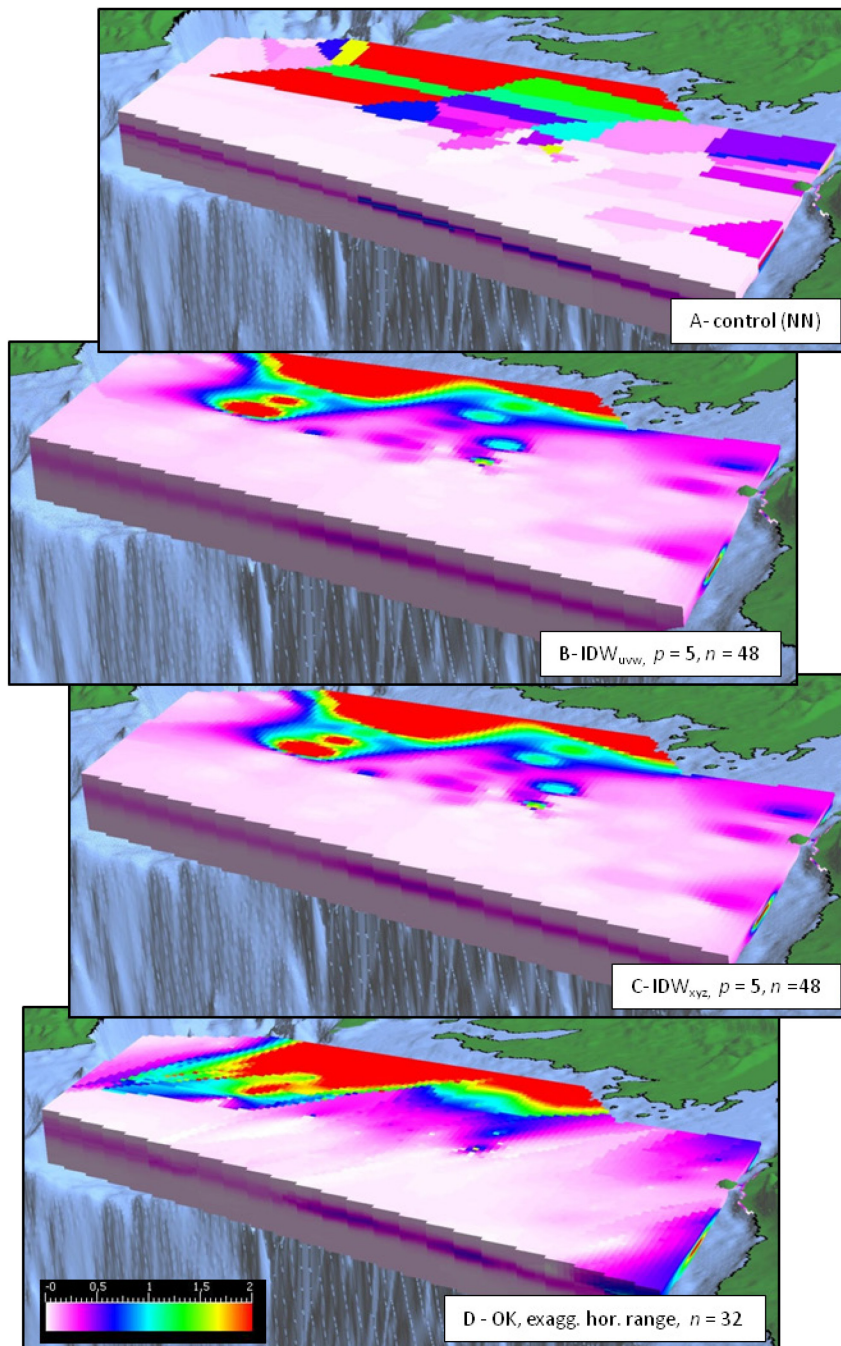


Figure 35. Représentations spatiales 3D de chl a (mg m³). Les 200 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200

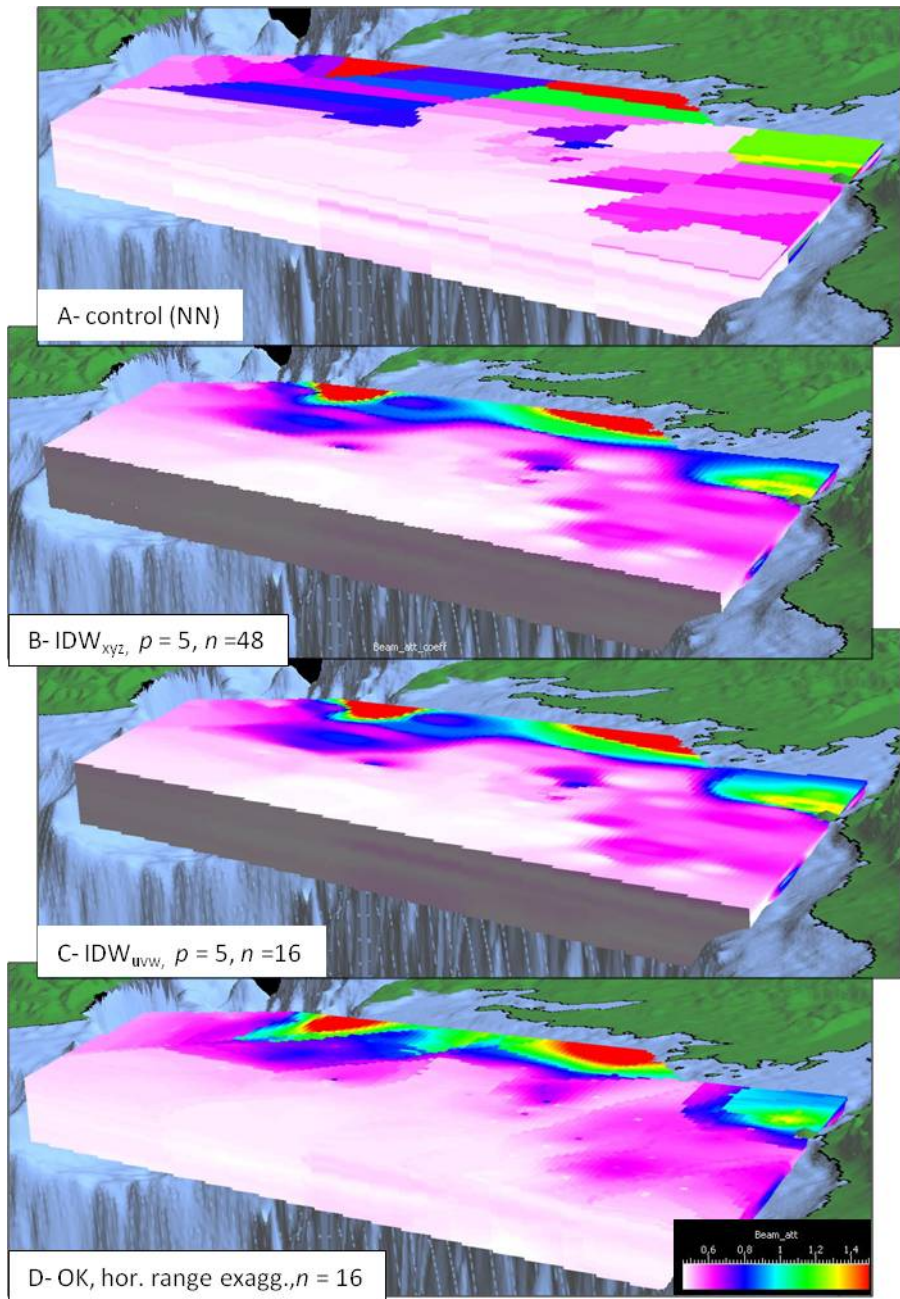


Figure 36. Représentations spatiales 3D de c_p (m^{-1}). Les 200 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 200

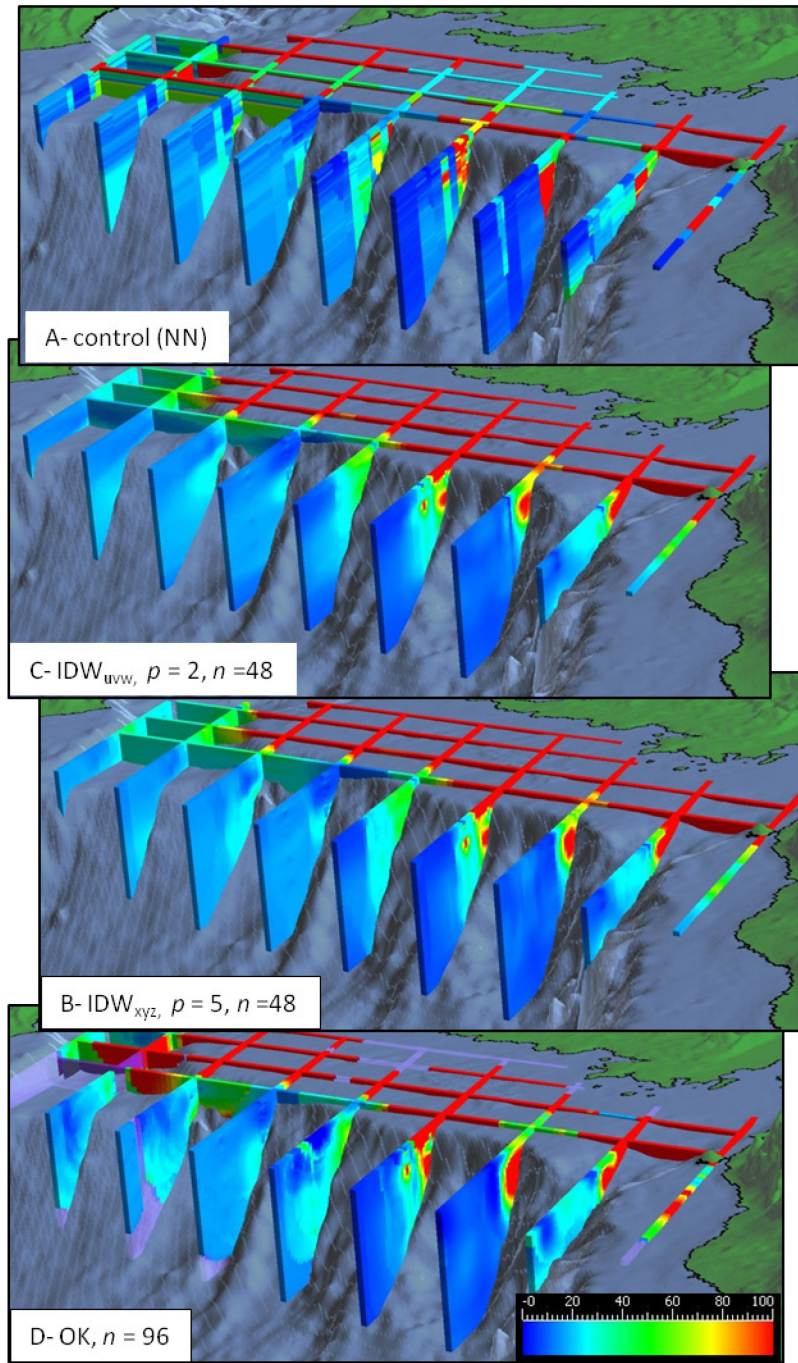


Figure 37. Représentations spatiales 3D du flux vertical de COP ($\text{mg C m}^{-2} \text{j}^{-2}$). Les 1700 m supérieurs de la colonne d'eau sont illustrés. L'exagération verticale = 50.