

Université de Montréal

L'utilisation d'ondes réfléchies pour cartographier l'écosystème  
benthique de Clambed, un site hydrothermal situé sur la dorsale

Juan de Fuca

par  
Sébastien Durand

Département des Sciences Biologiques  
Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de Maître ès sciences (M.Sc.)  
en biologie

Juillet, 2005

© Sébastien Durand, 2005.



d#

302

U54

2006

V.002

**Direction des bibliothèques**

**AVIS**

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

**NOTICE**

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé:

L'utilisation d'ondes réfléchies pour cartographier l'écosystème  
benthique de Clambed, un site hydrothermal situé sur la dorsale

Juan de Fuca

présenté par:

Sébastien Durand

a été évalué par un jury composé des personnes suivantes:

Daniel Boisclair  
président-rapporteur

Pierre Legendre  
directeur de recherche

S. Kim Juniper  
codirecteur

Denis Marcotte  
membre du jury

Mémoire accepté le

21 / 02 / 06

## RÉSUMÉ

Présentement nous utilisons des sous-marins, véhicules téléguidés, caméras, sonars et autres équipements de hautes technologies afin d'explorer et étudier des sources hydrothermales. Cachés au fond des océans, ces écosystèmes sont encore régis des dynamiques qui nous échappent. Sans l'aide de survol de haute définition décrivant de vastes étendues, les études spatio-temporelles n'ont aucune puissance.

À l'intérieur de ce projet, nous émettons l'hypothèse que les données récoltées par sonar sont polyvalentes et riches en information permettant l'identification et la cartographie de communautés, d'espèces et de types de substrats.

Parmi les divers aspects que nous abordons dans cette étude, nous touchons l'importance d'effectuer une bonne normalisation pour la profondeur ainsi que les différents types de variables à extraire des survols sonores. Nous avons différencié avec succès les cinq habitats dominants retrouvés à l'intérieur d'un champ hydrothermal en fonction de leurs signatures sonores et décrit ces dites signatures en fonction des variables que nous avons utilisées. Pour faire suite à cette étude pilote, afin d'identifier à quels éléments du paysage les signatures sonores sont sensibles, nous avons analysé les signatures sonores à l'aide d'une description exhaustive de tous les éléments visuels retrouvés dans les empreintes sonores. Apte à couvrir de vastes surfaces, cette technique est efficace et sensible aux différentes textures et densité associées aux communautés biologiques étudiées. Nous croyons que cette méthode est un moyen puissant permettant l'étude de la distribution de communautés, de leur évolution ainsi que des dynamiques des écosystèmes.

**MOTS CLÉS :** *Analyse canonique ; type d'habitat ; sources hydrothermales ; sous-marin téléguidé ; télédétection ; écho sonore ; vidéo.*

## ABSTRACT

Today, we use submarines, remotely operated vehicles, towed cameras, sonars, and many other high-tech instruments, to explore and study hydrothermal vents. In this world hidden under the sea, community dynamics needs to be understood. It is only through high-definition surveys carried out at broad scales that spatio-temporal studies become meaningful. In this project, we hypothesized that backscatter information would be a data-rich and versatile way of estimating and recognizing the distribution and composition of communities of species and substrate types, through the intensity signatures of their respective textures.

Throughout this research, many aspects of this remote sensing technique are discussed. For example, the importance of accurate depth normalisation and the types of variables that are to be extracted from such surveys. Tested on the five dominant habitats found in a hydrothermal vent field, using the described method, we successfully discriminated habitats based on their sonar signatures and also attempted to describe the sonar signatures types based on the many variables we used. After such initial testing, in an attempt to identify to what element of the landscape sonar signatures are sensible, we related the sonar signatures to an exhaustive video description made of the sonar footprint areas that had been sampled. This technique can be used in surveys covering broad areas because it is efficient and sensible to the textures and densities associated with the dominant biological communities encountered. We believe this method will greatly enhance studies of community distribution and evolution and of ecosystem dynamics.

KEY WORDS : *Canonical analysis; habitat types; hydrothermal vents; remotely operated vehicle; remote sensing; backscatter; video.*

## TABLE DES MATIÈRES

RÉSUMÉ . . . . .	iii
ABSTRACT . . . . .	iv
TABLE DES MATIÈRES . . . . .	v
LISTE DES TABLEAUX . . . . .	viii
LISTE DES FIGURES . . . . .	ix
LISTE DES SIGLES . . . . .	x
DÉDICACE . . . . .	xi
REMERCIEMENTS . . . . .	xii
 <b>CHAPITRE 1 : Introduction générale . . . . .</b>	 <b>1</b>
1.1 Les sources hydrothermales abyssales : origine, découverte et exploration	1
1.2 Les sources hydrothermales abyssales : les méthodes et leurs limites .	8
 <b>CHAPITRE 2 : The use of video surveys, a Geographic Information System and sonar backscatter data to study faunal community dynamics at Juan de Fuca Ridge hydro- thermal vents . . . . .</b>	     <b>12</b>
2.1 Introduction . . . . .	13
2.2 Material and methods . . . . .	15
2.2.1 Study sites . . . . .	15
2.2.2 Image and sonar acquisition . . . . .	15
2.2.3 Mapping and GIS construction . . . . .	15
2.2.4 Sonar backscatter experiments . . . . .	16
2.3 Results and discussion . . . . .	20

2.3.1	Overall trend . . . . .	20
2.3.2	Density increases and succession (1999-2000) . . . . .	21
2.3.3	Density decrease and senescence (2000-2001) . . . . .	21
2.3.4	Backscatter differentiation . . . . .	22
2.4	Conclusion . . . . .	26
2.5	Acknowledgements . . . . .	27

**CHAPITRE 3 : Sonar backscatter differentiation of dominant macro-habitat types in a hydrothermal vent field . . . . .**

3.1	Abstract . . . . .	29
3.2	Introduction . . . . .	31
3.3	Materials and methods . . . . .	34
3.3.1	Study site and habitat description . . . . .	34
3.3.2	Study design . . . . .	37
3.3.3	Signal processing methods . . . . .	38
3.4	Results . . . . .	46
3.4.1	Relationship between PCC and altitude . . . . .	49
3.5	Technical discussion . . . . .	52
3.5.1	The influence of altitude . . . . .	52
3.5.2	The nature of the five habitat type signatures . . . . .	52
3.6	Conclusion . . . . .	57
3.6.1	Technical implications for other domains . . . . .	59
3.7	Acknowledgments . . . . .	60

**CHAPITRE 4 : Classification et détection d'habitats benthiques à l'aide de signatures sonores . . . . .**

4.1	Résumé . . . . .	62
4.2	Introduction . . . . .	63
4.3	Méthodes . . . . .	65



	vii
4.3.1	Acquisition des données . . . . . 65
4.3.2	Traitement des données et extraction des variables . . . . . 65
4.4	Analyses et résultats . . . . . 68
4.4.1	Transects verticaux . . . . . 68
4.4.2	Transects horizontaux . . . . . 70
4.5	Discussion et conclusion . . . . . 72
<b>CHAPITRE 5 : Conclusion</b>	<b>. . . . . 73</b>
<b>BIBLIOGRAPHIE</b>	<b>. . . . . 77</b>
<b>APPENDICE A : The definitions of the VE variables along with their transformation details, the transformation trials for the VI data sets, and the variables retained by forward selection are shown.</b>	<b>. . . . . xiii</b>
<b>APPENDICE B : Habitat clusters, obtained from different altitude ranges using the SEL variable subset, and projected on the first and second discriminant axes.</b>	<b>. . . . . xx</b>
<b>APPENDICE C : Description and usage of the Sonarex package written in R language to analyse the Imagenex backscatters.</b>	<b>. . . . . xxviii</b>
C.1	sonarex . . . . . xxix
C.2	import.imatxt . . . . . xxxi
C.3	extract . . . . . xxxiv
C.4	analyse . . . . . xl
C.5	ldaTest . . . . . xlv
C.6	rdaTest . . . . . xlviii
C.7	graph.ldaTest . . . . . li
C.8	graph.rdaTest . . . . . lv

## LISTE DES TABLEAUX

2.I	Distribution percentages of polynoid and tubeworm assemblages . . .	17
2.II	Results of K-means partitioning of backscatter signatures for the 5 studied habitats . . . . .	24
3.I	Percentages of correct classification . . . . .	44
3.II	Habitat assignments of all echoes based on the VT data tables, using the complete variable sets (COM) . . . . .	51
3.III	Correlation between sonar variables and habitat types, for the three altitude ranges . . . . .	53
A.I	Definitions, transformations, and skewness values of the VE variables	xiv
A.II	Transformation trial for the VI data table and subtables . . . . .	xvi
A.III	Variables found in the FWD variable sets . . . . .	xviii

## LISTE DES FIGURES

2.1	Acoustic return area definition . . . . .	18
2.2	Map of Cloud vent site describing the successional changes between two faunal assemblages during the 1999-2000 period . . . . .	19
2.3	Example of contrasting habitat types and backscatter signatures at Clam Bed vent . . . . .	23
3.1	Habitats respective sonar signature and picture . . . . .	36
3.2	Echo segmentation of a real backscatter . . . . .	39
3.3	The overlaid intensities of every backscatters used . . . . .	41
3.4	Both original and power corrected backscatter profiles . . . . .	47
4.1	Discrimination d'échos acquis entre 7 et 10 mètres d'altitude . . . . .	69
4.2	Projection correlant 21 variables sonores sur deux axes canoniques . . . . .	71
B.1	Backscatter discrimination results using VI variables and subtable 1-4m . . . . .	xxii
B.2	Backscatter discrimination results using VI variables and subtable 4-7m . . . . .	xxiii
B.3	Backscatter discrimination results using VI variables and subtable 7-10m . . . . .	xxiv
B.4	Backscatter discrimination results using VE variables and subtable 1-4 m . . . . .	xxv
B.5	Backscatter discrimination results using VE variables and subtable 4-7 m . . . . .	xxvi
B.6	Backscatter discrimination results using VE variables and subtable 7-10 m . . . . .	xxvii

## LISTE DES SIGLES

CLAM	Habitat de palourdes
LIMP	Habitat de patelles
TUBE	Habitat de buisson de vestimentifaires
LAVA	Habitat de lave nue
PERI	Habitat en périphérie des buissons de vestimentifères
ROV	“Remotely operated vehicle” (ROV)
LBL	“Long base-line”
USBL	“Ultra-short base-line”
VE	Groupe de variables extraites
VI	Groupe de variables d’intensité
VT	Groupe de variables extraites et d’intensité

*«Dans l'honnêteté mordante du sel,  
la mer révèle ses secrets à ceux qui savent écouter.»*

Sandra Benitez

*«...il y a une mesure en toutes choses,  
et savoir la saisir à propos est la première des sciences.»*

Thémistocle

## REMERCIEMENTS

Il y a un gouffre entre le moment où l'on entame une maîtrise et le moment où le mémoire est enfin soumis. Le trajet d'un bord à l'autre est semé d'éprouvantes embûches et de nombreux détours. Je n'ai réussi à faire ce parcours que grâce à un certain nombre de personnes qui ont su me guider et me soutenir dans ce chemin non balisé.

Donc par ordre d'apparition :

Merci à Mike Fellow pour l'installation des paramètres de survol du sonar Imagenex 881 ; à Imagenex Corp. pour le support technique ; à Les J. Hamilton pour l'aide qu'il m'a apportée afin d'analyser mes échos ; à J. Illman pour sa grande expérience, ses blagues en mer et la formation de navigateur qu'il m'a donnée. Rencontré au cours des nombreuses expéditions en mer, qui m'ont bouleversées dans tous les sens du terme, je remercie l'équipage du navire de la garde côtière canadienne le John P. Tully, pour leur accueil et leur compétence. Plonger à 2200 mètres de profondeur est un exploit en soit, que j'ai vu réalisé à de nombreuses reprises par les membres de l'équipe du ROPOS. Je remercie donc tous les membres de l'équipe du ROPOS pour leur patience, leur humour et leur savoir-faire ainsi que Keith Shepherd pour les opportunités qu'il m'a données de découvrir l'océan et d'explorer le monde des abysses. Un gros merci à Stéphane Dray, un ami et collègue qui m'a initié au langage R et qui ma souvent guidé et sorti des mes incompréhensions statistiques ; à mon codirecteur Kim Juniper, qui m'a permis de vivre une expérience plus grande que nature dans deux océans du globe, qui a su me conseiller et répondre à mes questions portant sur l'écologie des sources hydrothermales. J'exprime ma gratitude à l'endroit de mon directeur de recherche Pierre Legendre, qui par la justesse de ses mots, la qualité de sa science et l'écoute, donc il a fait preuve, ma guidé et initié au travail et à la rigueur scientifique.

Enfin, je remercie les correcteurs anonymes de mes articles scientifiques pour leurs valeureuses remarques ainsi que tous les membres du comité-conseil pour leurs judicieuses recommandations et leur support ! Cette recherche a été parrainé par des subventions d'occasions de recherche concertée (ORC) du CRSNG accordé à S. K. Juniper and P. Legendre, et par "Canadian Scientific Submersible Facility".

Finalement, je remercie chaleureusement ma famille pour leur support inconditionnel et leur compréhension ; mes amis pour les nombreuses discussions, remises en question, et évasions psychologiques. Je remercie grandement ma copine Christine pour son amour et sa patience, qui tout au long de ce voyage a su être la boussole au creux de ma main et la muse chantant à mon oreille.

---

# CHAPITRE 1

## INTRODUCTION GÉNÉRALE

---

### 1.1 Les sources hydrothermales abyssales : origine, découverte et exploration

L'hydrothermalisme marin est un processus géothermal qui remonte à des périodes très reculées de l'histoire de la Terre. Il y a 4.4 milliards d'années, la Terre était à la fin de sa période d'accrétion. Devenue suffisamment massive et munie d'un champ gravitationnel apte à retenir les gaz, l'accumulation des émanations volcaniques a permis à la Terre de s'envelopper dans sa première atmosphère. Environ 400 millions d'années plus tard, une chute suffisante de la température a condensé la vapeur d'eau et sculpté les premiers nuages (Allegre and Schneider, 1994). Les averses diluviennes qui s'ensuivirent engendrèrent l'océan primitif, siège d'activité hydrothermale intense.

La compréhension des phénomènes reliés à ce genre d'activité terrestre naquit en 1915 avec Alfred Wegener qui développa, dans 'The Origin of Continents and Oceans', la première théorie sur la dérive des continents. En 1960, Harry Hess soutint la théorie de Wegener et expliqua le mécanisme permettant le déplacement des plaques tectoniques. Selon Hess, un petit nombre de plaques rigides mais mouvantes recouvrirait toute la surface de la Terre. Ces plaques seraient mues par les courants de chaleur internes de la Terre. Générées dans les zones de divergences par le reflux de magma, ces plaques seraient détruites dans les zones de subduction par leur fusionnement avec le manteau terrestre. Trois années plus tard, des inversions magnétiques enregistrées dans la croûte terrestre près de zones de divergence furent découvertes (Masson, 1958). Ce fût Fred Vine (Vine and Drummond, 1963) qui lia ce phénomène à la création de la croûte

terrestre dans les zones de divergence.

En 1977, la première source hydrothermale marine fut découverte par Corliss et Ballard au cours d'une plongée faite à 3000 mètres de profondeur le long des côtes des îles Galapagos (Ballard, 1977; Corliss and Ballard, 1977). Situées sur les zones de divergence des plaques tectoniques ou près de points chauds, les sources hydrothermales sont soutenues par l'activité géothermale de la Terre. Aux zones de divergence, l'eau de mer s'infiltré dans les fractures de la croûte océanique et forme de vastes cellules de convection qui peuvent atteindre de deux à quatre kilomètres de profondeur (Kelley et al., 2002; Wilcock et al., 2002). Pendant ce parcours, l'eau est acidifiée et est réchauffée par l'activité de la chambre magmatique. Elle se charge aussi de nombreux éléments par son contact constant avec la croûte océanique. Au niveau des événements hydrothermaux, on trouve dans ces fluides des sulfures métalliques, du méthane, de l'oxyde de carbone, du sulfure d'hydrogène et autres. Lors du contact de ces eaux chaudes avec l'eau de mer, ces minéraux dissous précipitent et forment des dépôts hydrothermaux extraordinairement riches en sulfures métalliques de fer, de zinc et de cuivre. Les teneurs minérales mesurées varient grandement et sont fonction de la température et de la force des flux hydrothermaux à leur point d'expulsion. Par exemple, les fumeurs noirs ont des températures de flux aux points d'expulsion variant entre 300 et 405 °C (Tivey et al., 1990). Ils rejettent, entre autres, du fer et de l'hydrogène sulfuré qui, se combinant dans l'eau, forment un panache noir rempli de particules de pyrite (FeS). Lorsqu'on les compare à celles des régions abyssales, les teneurs minérales expulsées par certains fumeurs noirs représentent un enrichissement par un facteur de  $10^8$ . Par contre, lorsque le fluide est dilué par de l'eau de mer avant son éjection, sa température chute et certains minéraux précipitent à l'intérieur des canalisations, laissant les minéraux plus solubles en solution. En conséquence, ce second type de fumeur, possédant un fluide variant entre 150 et 300 °C, produit en majorité du calcium, du baryum et du silicium. Ces minéraux précipitent lors du contact avec l'eau de mer et forment un nuage blanc composé de sulfures, tel que l'anhydrite soluble ( $\text{CaSO}_4$ ) (Kelley et al., 2002).



À des profondeurs de 2000 à 4000 mètres, la noirceur absolue règne sur des eaux avoisinant les 2 °C. Les organismes des plaines abyssales, adaptés à des pressions de 300 atmosphères, d'une densité moyenne de vingt vers et ophiures par mètre carré (Tunnicliffe and Thomson, 1999), ne subsistent que grâce aux maigres ressources alimentaires provenant de la couche photique. Cependant dans le milieu voisin, à proximité des sources hydrothermales, la présence de communautés d'organismes uniques prospérant dans un écosystème riche a bouleversée le monde de la biologie. Comparativement aux écosystèmes de surface où la chaîne alimentaire dépend de la photosynthèse, les sources hydrothermales furent le premier exemple d'écosystème fondé exclusivement sur la chimiosynthèse à être découvert. La photolyse rendue impossible par l'absence de lumière, l'échelon primaire de cet écosystème utilise l'énorme potentiel énergétique dérivé des fortes concentrations de composés chimiques et de la chaleur élevée des fluides. Cette flore bactérienne, composée de bactéries autotrophes, oxyde des minéraux réduits pour produire l'énergie chimique nécessaire au métabolisme. Retrouvées non seulement en suspension, sur les substrats et dans les réseaux des canalisations hydrothermales sub-superficiels (Jannasch and Mottl, 1985), mais aussi sur et à l'intérieur même de certains organismes, ces bactéries jouent un rôle capital. La biomasse associée aux sources hydrothermales peut-être de 500 à 1000 fois supérieures à celle des déserts abyssaux (Juniper, 2003). En périphérie de ces grands fumeurs noirs près des flux diffus, l'on retrouve des communautés constituées de petits organismes d'une grande diversité taxinomique et d'une densité pouvant atteindre un demi-million d'animaux par mètre carré (Tunnicliffe and Thomson, 1999).

Hormis le fait que ce nouvel écosystème soit très productif, il est fortement instable et précaire. Un champ hydrothermal ne possède qu'une longévité de quelques centaines d'années et les cheminées elles-mêmes, vestiges de l'activité thermique de ces sites, ne sont maintenues actives que de 10 à 20 ans (Tunnicliffe, 1990). En oxydant le sulfure d'hydrogène, les bactéries produisent l'énergie nécessaire pour fixer le carbone du CO<sub>2</sub> retrouvé dans l'eau. C'est donc à partir de cette source de carbone organique

que l'étrange chaîne biologique de cet écosystème s'est développée. La première liste d'animaux des sources hydrothermales, publiée en 1985, montrait des organismes dont la plupart étaient inconnus de la science ; de nouvelles espèces, des genres et même, des familles furent créées. La faune qui habite ce milieu est principalement composée d'invertébrés tels que des vers, des mollusques et des arthropodes ; cela est étonnant puisque partout ailleurs sur les fonds océaniques, ce sont les éponges, les anémones et autres qui dominent (Grassle, 1986). En fait, 93% des 375 espèces retrouvées auprès des sources ne vivent qu'au voisinage des fluides hydrothermaux, ce qui représente un taux d'endémisme exceptionnel en milieu océanique (Tunnicliffe and Fowler, 1996). Comparativement au reste de l'océan profond, les communautés hydrothermales comportent des espèces animales (mollusques et crabes) qui ont conservé des caractères primitifs connus seulement chez les espèces fossiles. Ces faits combinés et décrits dans Tunnicliffe (1992), suggèrent que cet écosystème possède une faune d'une origine ancienne datant du Paléozoïque qui aurait suivi une évolution biogéographique associée depuis fort longtemps aux événements hydrothermaux.

Depuis cette découverte d'importance, plusieurs chercheurs s'interrogent sur le rôle joué par ce milieu extrême dans l'apparition de la vie. Selon une hypothèse généralement acceptée, mais mise en doute par (Brasier et al., 2004), les stromatolites datant de 3.460 milliards d'années seraient les plus vieux fossiles actuellement connus. Ils auraient été formés par l'activité de micro-organismes photosynthétiques marins similaires aux cyanobactéries (Schopf, 1999). Parmi cet ancien groupe d'organismes que sont les procaryotes, les archéobactéries représentent la forme de vie la plus primitive sur notre planète, mais aussi la plus extrémophile, puisqu'on y trouve des organismes thermoacidophiles, hyperthermophiles, halophiles extrêmes et méthanogènes. Encore aujourd'hui, à l'intérieur des murs des cheminées hydrothermales, on retrouve une des archéobactéries les plus tolérantes à la chaleur : *Pyrolobus fumarii* se reproduit à 113 °C (Blöchl et al., 1997; Kashefi and Lovley, 2003). En examinant attentivement les conditions physiques et chimiques extrêmes (telles que fortes températures, concentrations chimiques

et pressions) présentes dans le réseau subsuperficiel des sources hydrothermales, on remarque que ces sites conservent encore de nos jours des conditions similaires sinon identiques à celles de l'océan archéen (Nisbet and Sleep, 2001). Cela fait de cet environnement un milieu à tout le moins propice au maintien d'organismes anciens, comme les *Methanopyrus*, un genre d'archéobactérie méthanogène hyperthermophile vivant aussi dans les cheminées : on les situe près de la base de l'arbre séparant les archéobactéries des bactéries et des eucaryotes. Bref, pour l'observateur attentif, les sources hydrothermales forment des oasis qui semblent ouvrir une fenêtre permettant un regard sur un Ancien Monde qui a su échapper au temps.

De façon plus descriptive, les peuplements hydrothermaux renferment de multiples microhabitats auxquels différents organismes se sont spécifiquement associés. Du centre à la périphérie de ces zones, les températures, les concentrations chimiques et la quantité de ressources nutritives diminuent avec la distance. Vues de haut, ces communautés d'organismes forment des zones concentriques autour des points de rejet. Par exemple *Paralvinella sulfincola* est un annélide qui est souvent la première espèce à coloniser le substrat des nouvelles cheminées et qui doit subir les conditions environnementales les plus extrêmes de cet écosystème, c'est-à-dire une pluie constante de précipités, de fortes concentrations de composés toxiques ( $H_2S$ ) et de très fortes températures (jusqu'à  $105^{\circ}C$ ). Là où les conditions sont moins extrêmes, certaines espèces ont développé des relations symbiotiques avec les bactéries autotrophes afin de subvenir à leurs besoins nutritifs. Les vestimentifères sont des vers tubicoles qui se nourrissent uniquement des bactéries symbiotiques situées à l'intérieur d'un organe interne appelé le trophosome. Chez les Vesicomylidae *Calymene magnifica* et *C. pacifica*, qui sont des palourdes, on a retrouvé des bactéries symbiotiques localisées dans leurs branchies hypertrophiées (Kochevar and Childress, 1996). Cette adaptation, qui répond entièrement aux besoins nutritionnels de l'organisme hôte, génère une forte dépendance aux composés retrouvés dans les fluides thermaux tels que l'HS et le  $CO_2$ . Le fait d'avoir une relation si directe avec les bactéries permet à ces organismes de proliférer rapidement et d'établir des

communautés denses et florissantes. Chez d'autres espèces, différentes stratégies sont utilisées. *Bathymodiolus thermophilus*, une autre espèce de bivalve, utilise trois modes d'alimentation : elle se nourrit de ses symbiotes, elle récolte des particules en suspension et elle absorbe de petites particules organiques (Nelson et al., 1995). Grâce à cette diète mixotrophe, cette espèce est souvent retrouvée à la périphérie des sources.

Les conditions physiques et chimiques extrêmes des champs hydrothermaux créent une barrière écologique qui protège les peuplements qui y sont adaptés des prédateurs provenant d'autres écosystèmes. Les colonies de vestimentifères forment des buissons qui offrent protection et nourriture à d'autres espèces. D'un point de vue écologique, ces buissons forment le centre des communautés animales qu'on retrouve aux sources hydrothermales de la dorsale Juan de Fuca, car une grande partie de la diversité biologique leur est associée. On y trouve des poulpes, des patelles, des palourdes, des crabes, des poissons et de nombreuses espèces de ver. Au sommet de la chaîne alimentaire, trois groupes de carnivores dominent. Chez les mollusques, on trouve par exemple des bigorneaux, des pieuvres, des poulpes blancs et des poulpes à oreilles. Chez les crustacés, *Bythograea thermydon* est un petit crabe blanc aveugle qui abonde dans les buissons de vestimentifères. Un peu plus gros, *Cyanograea praedator* est un crabe des fumeurs qui se nourrit de polychètes tels que *Paralvinella sulfincola* et d'autres espèces retrouvées à proximité des fumeurs. Ces deux espèces de crabe n'ont jamais été observées en dehors des sites hydrothermaux, mais un de leurs cousins, le crabe-araignée (*Macroregonia macrochira*), est une espèce beaucoup plus grosse et non endémique ; c'est un des grands prédateurs de cet écosystème. Dans ce rôle de prédateur, il est accompagné par diverses espèces de poissons comme le *Thermarces cerbrus*, un poisson blanchâtre de 30 cm de longueur ressemblant à une anguille obèse ou de grosseur plus imposante : le grenadier (*Coryphaenoides acrolepis*), le chabot à grosse tête (*Psychrolutes phrictus*) ainsi que d'autres espèces visitent cet écosystème riche.

Depuis la découverte des sources hydrothermales, les scientifiques développent de nouveaux outils et de nouvelles technologies afin de percer les secrets de cet écosys-

tème datant des premiers âges de la Terre. Les scientifiques ont d'abord utilisé des sous-marins habités pour explorer cet environnement reclus. En 1989 le premier ROV ("Remotely Operated Vehicle") fut utilisé pour affranchir la recherche des besoins et limites des équipages humains. Cela a marqué le début d'une nouvelle ère d'exploration. Dorénavant, l'étude des sources se fait à l'aide de robots, véhicules téléguidés, caméras, sonars et plusieurs autres instruments de haute technologie.

## 1.2 Les sources hydrothermales abyssales : les méthodes et leurs limites

En pleine mer, étudier un écosystème situé à deux ou trois mille mètres de profondeur est une tâche complexe. Depuis la découverte des sources hydrothermales en 1977, la façon de les d'étudier n'a guère changé. Au fond de l'océan, les techniques vidéo et photographiques jouent un rôle clef. Ces instruments servent non seulement à la navigation et à l'exploration des sites de plongée, mais ils permettent également aux scientifiques d'échantillonner et d'effectuer des études taxinomiques, écologiques et comportementales.

Les méthodes vidéo sont très informatives, mais elles dépendent de la lumière. À proximité des sources hydrothermales, l'éclairage fourni par les sous-marins et les véhicules téléguidés est souvent insuffisant. La lumière est très vite absorbée par l'eau et par les nombreuses particules minérales et bactériennes mises en suspension par les flux hydrothermaux. Ce phénomène limite l'obtention d'enregistrements vidéo de qualité à des survols effectués à de très basses altitudes (près du fond). Parmi les champs de roches basaltiques se dressent des cheminées aux formes irrégulières et tranchantes. Façonnées par l'activité des sources hydrothermales, ces cheminées sont structurellement instables et lorsqu'elles sont actives peuvent émettre des fluides dépassant les 400 °C. Lors de survols effectués à basse altitude, les pilotes doivent redoubler de vigilance afin de prendre en considération tous ces facteurs, ce qui a pour effet de rendre de telles manoeuvres très dangereuses et astreignantes.

Malgré tout, de nombreuses cartes et mosaïques de photos ont été construites à partir de méthodes vidéo automatisées. Malheureusement, les données issues de ces méthodes sont difficiles à traiter et prennent un temps considérable à extraire. Le problème tient au fait que les images utilisées doivent être étirées et corrigées en fonction du relief survolé ; trois ou quatre points lasers, projetés depuis le submersible, fournissent le référentiel. En pratique, même pour des survols vidéo réussis effectués entre trois et cinq mètres au-

dessus du fond, les grands taux d'absorption et la résolution du support vidéo rendent périlleuse l'identification des organismes ; il en va de même pour les points lumineux des lasers qui sont souvent à peine visibles. Dans le cadre de ce mémoire, l'identification minutieuse des organismes et des types de substrats vus pour chacune des secondes d'un survol vidéo de trois heures a pris plus de deux mois à être effectué. En dépit des efforts déployés et la faible distance séparant les organismes de l'objectif de la caméra, certaines classes d'organismes n'ont pu être identifiées par manque de visibilité et de résolution. En conséquence, il devient difficile d'obtenir des images claires de la réalité spatio-temporelle des sites étudiés à l'aide de telles méthodes. Considérant les coûts élevés associés aux expéditions en mer et la praticabilité réduite de ces techniques, les scientifiques sont forcés de se tourner vers d'autres méthodes pour les études de la structure spatiale de ces écosystèmes.

Contrairement aux méthodes utilisant la lumière, les méthodes sonores ont l'avantage d'utiliser des ondes qui se propagent très bien en milieu aqueux et qui sont indépendantes des sources lumineuses. Alors que certains mammifères marins sont passés maîtres en la matière, l'utilisation du son à des fins de télédétection est un outil encore peu maîtrisé par l'homme. Le premier sonar passif (écoute seulement, pas d'émission), créé par Lewis Nixon en 1906, fut construit pour permettre la détection des icebergs. Après les modifications déterminantes de Paul Langevin, qui étaient fondées sur les propriétés piézo-électriques du quartz, le sonar passif a permis la détection des sous-marins au cours de la Première Guerre mondiale. En 1918, conjuguant récepteur et émetteur, l'Angleterre et les États-Unis ont développé les premiers sonars actifs. Au cours de la Deuxième Guerre mondiale, les Américains ont créé l'acronyme "sonar" à partir de l'expression "Sound Navigation and Ranging". Plusieurs années plus tard, en 1972, R. H. Belderson publie "Sonographs of the seafloor", un des premiers livres décrivant clairement les différentes étapes associées à l'interprétation de signaux sonores permettant la création de cartes bathymétriques. Afin d'accomplir cette tâche, en réponse aux multiples avancements technologiques, divers types de sonars sont maintenant dispo-

nibles sur le marché. Il en existe trois grandes catégories : les échosondeurs, les sonars à balayage latéral et les sonars à multifaisceaux. La distinction principale entre ces divers modèles est le nombre de faisceaux sonores utilisés. L'augmentation du nombre de faisceaux donne une meilleure résolution et une plus grande couverture. Contrairement aux deux groupes possédant plusieurs faisceaux (généralement deux pour les sonars à balayage latéral et des dizaines pour les sonars à multifaisceaux), l'échosondeur, lui, en utilise un seul.

Dans le cadre de ce mémoire, nous présentons une série de développements méthodologiques visant à améliorer la cartographie écologique des fonds marins à l'aide de données sonores. Nous tentons d'identifier et d'extraire des signatures sonores à partir des échos enregistrés. Ces signatures sonores nous permettront d'identifier les habitats dominants retrouvés dans un champ hydrothermal.

Pour obtenir des informations biologiques et environnementales pertinentes, l'étude doit porter sur les petites aires de distribution des habitats qu'on retrouve à l'intérieur d'un champ hydrothermal. Les sonars à multifaisceaux ou latéraux ont été conçus pour d'échantillonner de grandes surfaces. Lorsqu'on utilise des faisceaux multiples, l'écho enregistré correspond à la somme des différentes réflexions produites ; par conséquent, l'écho enregistré ne peut pas être redivisé. Ainsi, l'information contenue dans cet écho peut seulement être utilisée pour caractériser l'ensemble de la zone échantillonnée. L'acquisition d'échos aptes à décrire de petites cibles ne peut se faire qu'avec un échosondeur.

Ce mémoire de maîtrise est composé de trois articles qui se recoupent légèrement. Le premier article, publié en 2002, présente deux méthodes d'investigation basées respectivement sur l'information visuelle et sonore. Nous avons utilisé des survols vidéo acquis au cours de trois années consécutives pour étudier la dynamique des communautés fauniques d'un champ hydrothermal. Après avoir dessiné les cartes annuelles de distribution d'habitat à partir de ces survols vidéo, nous les avons comparées et analysées à l'aide du logiciel Idrisi, un système d'information géographique (SIG). Nous



avons évalué les différentes limitations associées au survol vidéo. Les résultats préliminaires d'une étude pilote sont également présentés ; cette étude a pour but d'évaluer la capacité qu'ont les signaux sonores de différencier les habitats en fonction de leurs différences de texture et de densité.

Constatant les résultats encourageants de cette étude pilote, le deuxième article, qui est en ce moment sous presse, approfondit divers aspects associés à la méthode de télédétection par sonar. Après avoir décrit de façon détaillée notre méthode d'analyse, nous testons si l'information extraite des signaux sonores de haute fréquence permet la discrimination précise des cinq habitats dominants que nous avons retrouvés dans le champ hydrothermal à l'étude. Nous évaluons aussi l'influence de la taille des empreintes sonores sur la qualité des signaux enregistrés. Puis, à titre exploratoire, nous utilisons et testons divers jeux de variables afin d'identifier celles qui sont les plus sensibles aux divers types de texture et de densité associées aux habitats échantillonnés par le sonar. Des annexes de l'article décrivent les détails de la méthode. Une bibliothèque de programmes en langage statistique R permettra à la communauté scientifique d'utiliser les méthodes de décomposition des signaux sonores que nous avons développées de même que les méthodes d'analyse de données que nous avons programmées et utilisées dans cet article.

Le dernier article, publié en 2005, traite plus en détail des aspects méthodologiques et statistiques reliés à la télédétection sonore. Le jeu de données présenté dans le deuxième article est constitué de transects verticaux ; nous l'utilisons ici pour explorer et évaluer le potentiel de discrimination de notre méthode, et ce, pour différentes tailles d'empreintes sonores. Finalement, nous utilisons un nouveau jeu de données constitué de multiples transects horizontaux pour étudier les corrélations reliant les divers éléments du paysage aux variables extraites des signaux sonores enregistrés.

---

## CHAPITRE 2


# THE USE OF VIDEO SURVEYS, A GEOGRAPHIC INFORMATION SYSTEM AND SONAR BACKSCATTER DATA TO STUDY FAUNAL COMMUNITY DYNAMICS AT JUAN DE FUCA RIDGE HYDROTHERMAL VENTS

---

Sébastien DURAND<sup>1,2</sup>, Marlène LE BEL<sup>1</sup>, S. Kim JUNIPER<sup>1</sup>  
and Pierre LEGENDRE<sup>2</sup>

Cahiers de Biologie Marine. (2002) 43 : 235-240.

<sup>1</sup> GEOTOP - Université du Québec à Montréal, C.P. 8888 succ.  
Centre-Ville-Montréal (Québec), H3C-3P8, Canada <sup>2</sup> Département des Sciences  
Biologiques - Université de Montréal, C.P. 6128 succ.  
Centre-Ville-Montréal (Québec), H3C-3J7, Canada



## 2.1 Introduction

The study of benthic faunal dynamics in the deep sea is often hampered by technical limitations that include our inability to accurately reconstruct or view areas of seafloor larger than what can be seen through submersible view ports or camera lenses. Manual mapping and image mosaicking have been used with some success to analyse organism distribution within vent sites and at the scale of hydrothermal fields (Juniper et al., 1998). In both cases, video or photo transects serve as the primary source of information. Image mosaicking, in principle, is less subjective than mapping by hand, but the present state of the art requires precise control of camera distance and viewing angle to avoid spatial distortion as images are adjusted and fitted into the mosaic (Grehan and Juniper, 1996). The reality of operating submersibles over the rugged volcanic and sulphide edifice terrains where hydrothermal vents are found does not always permit collection of imagery suitable for mosaicking, and so manual mapping from video records is still used in some applications (Sarrazin et al., 1997). Both techniques yield useful maps that can be used to elucidate relationships between vent organism distribution and habitat features. Time-series mapping with either method can be used to study organism growth, faunal succession, and the response of species and communities to environmental change. Cartographic studies of vent fauna can be further developed by integration of map data into a Geographic Information System (GIS), where a range of powerful, automated spatial analysis tools can be used to explore ecological questions.

This paper reports on a combined hand mapping and GIS approach that we have been using to study early faunal community dynamics at Cloud Vent, a new hydrothermal site created by a January 1998 eruption on Axial Volcano, on the Juan de Fuca Ridge. We also report on preliminary results of organism and substratum differentiation in sonar backscatter signals, and consider their complementarity with visual imagery. Sonar backscatter analysis offers a potential for automated, swath mapping of habitat over large areas of seafloor, at scales that go well beyond what can be practicably

achieved with visual imagery. Seafloor geological and biological features influence the strength and temporal distribution of the acoustic signal that is reflected back to the detector (the first echo), according to their respective densities and textures.

## 2.2 Material and methods

### 2.2.1 Study sites

Cloud Vent ( $45^{\circ}56'00''\text{N}$  -  $129^{\circ}58'53''\text{W}$ ) is located on the South-East of Axial Volcano caldera ( $46^{\circ}\text{N}$  -  $130^{\circ}\text{W}$ ) on the Juan de Fuca Ridge, 200 nautical miles off the British-Columbia coast. The site is 1500 m deep and covers a 12m x 10m area of shallow collapse in the central portion of the lava flow generated by the January 1998 eruption. This site was the focus of our video transects and GIS mapping study. Sonar backscatter experiments were conducted at Clam Bed ( $47^{\circ}57'47''\text{N}$  -  $129^{\circ}05'30''\text{W}$ ), a larger vent field located on the Endeavour Segment of the Juan de Fuca Ridge. The *circa* 50m x 20m site consists of two roughly parallel 2-3m high north-south trending ridges colonised by tubeworm (*Ridgea piscesae*) and alvinellid polychaete/limpet assemblages. The depression between the ridges is slightly sedimented and hosts vesicomid clams.

### 2.2.2 Image and sonar acquisition

The Remotely Operated Vehicle ROPOS (ROV) was used for acquisition of all video and sonar imagery. Imagery from the submersible's SIT (Silicon Intensified Targeting) and 3-CCD (Charge-Coupled Device) cameras was recorded to S-VHS or digital (Mini-DV) tapes. Video imagery was recorded during overlapping zigzag imaging transects at Cloud Vent, where the submersible held a constant heading. The SIT video imagery served to determine the general site profile and to reconstruct geological features. The 3-CCD colour camera was used for faunal community identification and characterisation.

### 2.2.3 Mapping and GIS construction

Vent faunal distribution maps for 1998, 1999, 2000 and 2001 were constructed following the mapping methods described in (Delaney et al., 1992; Sarrazin and Juniper, 1998;

Sarrazin et al., 1997). Briefly, while repeatedly viewing the recorded video transects, visually distinct faunal assemblages were hand drawn on a background geological map that included two reference markers (N4 & N6). Information on the relative density of the faunal assemblages was also incorporated into the distribution maps. Hand drawn maps were then scanned, and imported into a graphics program (Freehand®) where the distribution of each faunal assemblage was traced into a separate layer, along with relative density information. Each layer was first imported into the GIS as separate a georeferenced layer, and then integrated into a single GIS using Idrisi® software. Each layer was georeferenced using the positions of the visible markers, and transect beginning and end positions, as determined by the submersible's long baseline navigation system (using the Universal Transverse Mercator, Zone 9 North "UTM-9N" coordinate system). As a result of the resampling or "georeferencing" process, the initial image files were stretched and rotated in order to fit the mapping projection (UTM-9N), thus generating directly comparable maps.

#### **2.2.4 Sonar backscatter experiments**

Sonar backscatter investigations used an Imagenex 881 mechanical scanning sonar, which has a resolution of  $0.15^\circ$ , a pencil beam of  $1.7^\circ$ , a frequency of 675 kHz and a maximum range of 100 m. In order to evaluate the ability of the sonar to distinguish between different bottom types and different benthic communities, we conducted a series of ground truthing experiments. The submersible was positioned over 5 visually distinct habitat or bottom types in the Clam Bed field, and sonar data were collected as the vehicle slowly rose vertically from 0m to 25m altitude while insonifying the seafloor below. For these experiments, the sonar head was fixed stationary and facing straight down (zero degree angle). The selected habitat/bottom types were:

- 1) Ridge top without tubeworms
- 2) Ridge top with semi-continuous tubeworm bushes

- 3) Ridge top with dense, continuous tubeworm bushes
- 4) Ridge top with polychaete/limpet assemblages
- 5) Sedimented depression with clams

At least 40 backscatters for each habitat type, obtained between 2.5 and 3.5m altitude, were included in this preliminary analysis. At this height, the footprint (insonified area) is a 9 cm diameter circle. Once the sonar data were post-processed (filtered for ambient noise, then normalized for depth following Clarke and Hamilton, 1999), the backscatter tail was extracted from the first echo (Fig. 2.1) and intensity/time values were analysed using Principal Components Analysis (PCA). The principal components that accounted for 95% of the variation of the original multivariate data were retained for partitioning. K-means partitioning (Legendre et al., 2002) was then performed on the principal components (coordinates of the objects in PCA ordination space) to group the different signature types. The best partition or grouping produced was selected using the Calinski-Harabasz (C-H) criterion, which is a  $F$ -statistic.

Table 2.I: Yearly distribution of polynoid polychaete and tubeworm assemblages within the two mapped zones (N4 and N6) at Cloud Vent, Axial Volcano. Data are percentages of the area in each zone occupied by the two assemblages.

Site (m <sup>2</sup> )	N4 (11.56)			N6 (34.76)			Total area (46.32)		
	1999	2000	2001	1999	2000	2001	1999	2000	2001
Polynoids (%)	29.33	37.55	13.02	10.14	6.75	1.81	14.93	14.44	4.61
Tubeworms (%)	0.66	7.05	5.05	3.60	27.31	14.55	2.87	22.25	12.18
Total area (%)	29.99	44.59	18.06	13.74	34.07	16.36	17.80	36.69	16.79

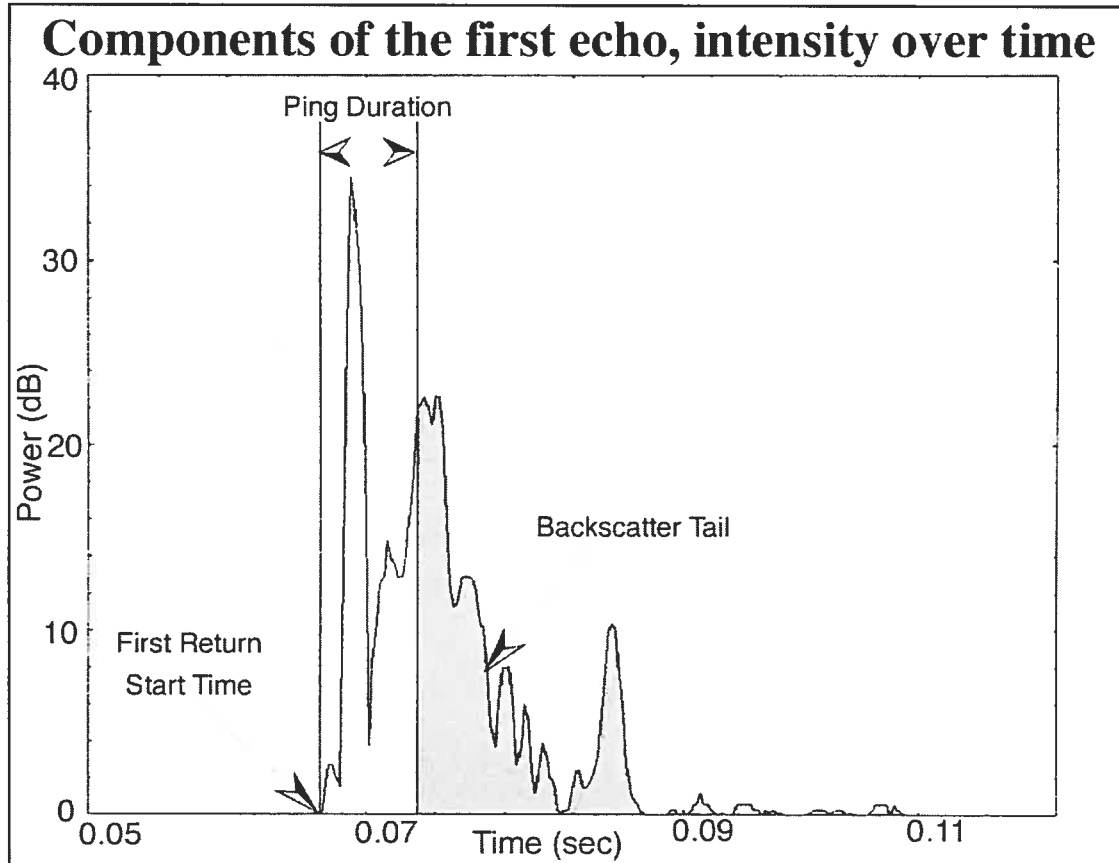


Figure 2.1: Acoustic return (signal strength versus time) detected by the Imagenex 881 sonar head following initiation of a single ping. The backscatter tail (shaded area) is the entire first echo minus the duration of the sonar interrogation ping. The shape of the signal (power fluctuation over time) is a function of bottom texture and density.



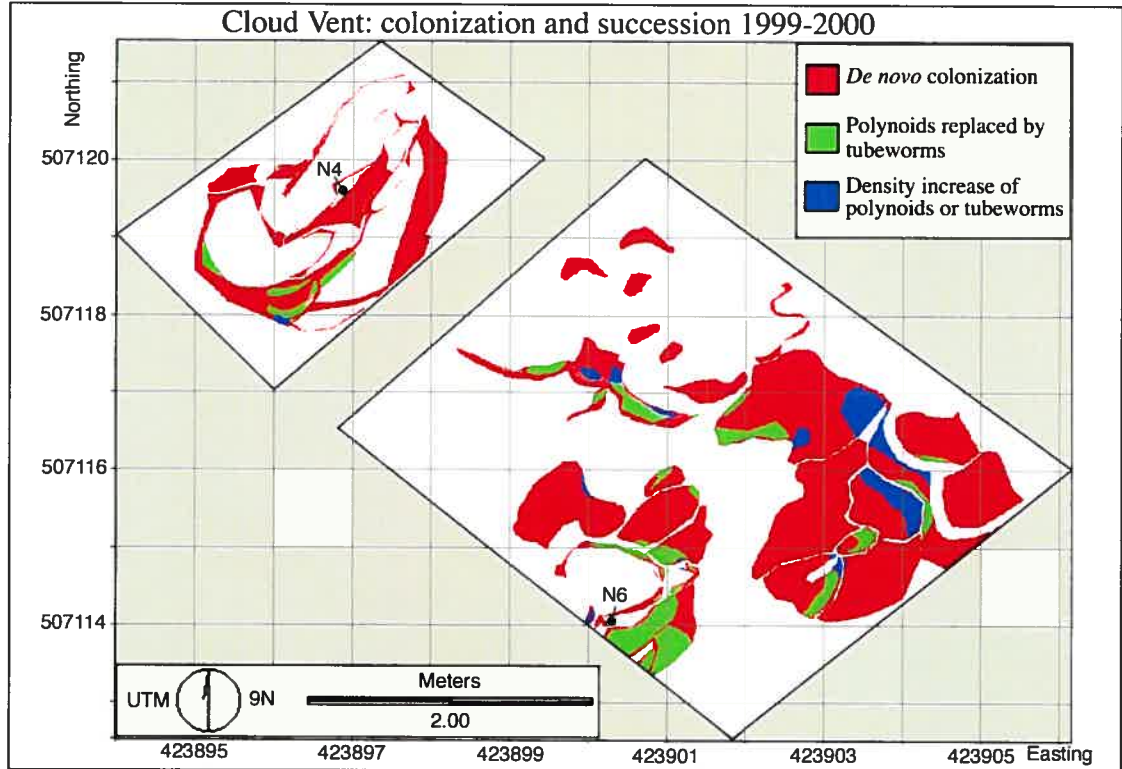


Figure 2.2: Map of Cloud Vent showing the zones where polynoids and tubeworms increased in density, as well as the successional changes between these two faunal assemblages during the 1999-2000 period. 1.6% of this area showed simple density increases of these two communities over the 1-year interval, while on 3.6% of the surface polynoids were replaced by tubeworms, and 24.0% of the study area was colonised *de novo* by tubeworms.

## 2.3 Results and discussion

Poor visibility in some years limited visual mapping to separate zones around the two markers placed in 1998. The zones were estimated to be 12.6m<sup>2</sup> (marker N4) and 36.8m<sup>2</sup> (marker N6) in area. Two visually distinct faunal assemblages were identified and mapped during the 4 years of the study. Although the composition of the assemblages has not been formally described, we designated them as the polynoid polychaete and tubeworm assemblages. The polynoid assemblage, which was the first to appear (1999), was dominated by the polynoid polychaete *Branchinotogluma sp.*. The tubeworm assemblage, which did not appear until the 2000 was dominated by the vestimentiferan *Ridgea piscesae*, the alvinellid polychaete *Paralvinella pandorae*, *Branchinotogluma sp.* and the limpet *Lepetodrilus fucensis* (J. Marcus, University of Victoria, personal communication).

Once the faunal distribution information was incorporated into the GIS, different analytical software tools were used to study the spatial and temporal relationships between the species, and between species and their environment. Table 2.I summarizes the year-to-year distribution of the polynoid and tubeworm assemblages in the mapped zones around markers N4 and N6.

### 2.3.1 Overall trend

In the summer of 1998, approximately 8 months after the eruption that created Cloud Vent, there were no visible faunal assemblages in the zones around the two markers. The following year, the visible macrofauna in both zones consisted almost entirely of polynoid polychaetes, with a few rare patches of tubeworms (Table 2.I). By 2000, the faunal compositions of zones N4 and N6 had begun to diverge. At N4, more substratum was still occupied by polynoid assemblages than by tubeworms, while at N6 tubeworms became the dominant group. In the final year of the study, there was

no noticeable shift in dominant groups in either zone but there was a general decrease in areal occupation by all faunal assemblages. This observation is compatible with an observed overall decrease in hydrothermal outflow at this site (S. Durand & M. Le Bel. personal observation).

### 2.3.2 Density increases and succession (1999-2000)

We undertook a detailed analysis of the progression of faunal colonization in the 1999-2000 interval, during which the most significant biological changes occurred at the site. In this analysis, we asked three questions of the database: 1) Did further *de novo* colonization of previously bare surfaces occur during this interval? 2) To what extent was the initial polynoid polychaete assemblage replaced by vestimentiferan tubeworms? 3) Were there areas where the initial (mostly polynoid) assemblage remained dominant and increased in density?

The results of these analyses are summarised in Figure 2.2. The dominant change during the 1999-2000 interval was *de novo* colonization of bare surfaces by tubeworm assemblages. We found that 24.0% of the study area was colonised *de novo* by tubeworms between 1999 and 2000, while only 3.6% of the surface occupied by polynoid assemblages was taken over by tubeworms. A further 1.6% of the total analysed area simply showed density increases for the assemblage present in 1999.

### 2.3.3 Density decrease and senescence (2000-2001)

We undertook a second interrogation of the database in order to evaluate the consequences of a noticeable decrease of hydrothermal activity in the 2000-2001 interval. Our initial questions were similar to the above consideration of density increases and succession: 1) To what extent did existing communities simply decrease in density? (2.4% of the colonised surface), 2) To what extent were tubeworms replaced by poly-

noids (8.4% of colonised surface) 3) To what extent did the fauna disappear altogether (20.8% of the total faunal colonised surface)?

The greatest change in the 2000-2001 interval involved the complete disappearance of existing faunal assemblages (not shown). Slightly more than 20% of the previously colonised surface was bare in 2001. On a further 8.4% of the colonised surface, tubeworm assemblages were replaced by polynoids. With respect to the latter point, we note that the dominant species in the polynoid assemblage (*Branchinotogluma sp.*) was also present and abundant in the tubeworm assemblage. Thus the “replacement” likely involved, to a large extent, the disappearance of the tubeworm overstory, as well as the species living on their tubes, to expose the polynoids on the substratum. Finally, for another 2.4% of the surface colonized in 2001, there was a decrease in density of the same faunal assemblage mapped in 2000.

### 2.3.4 Backscatter differentiation

In the K-means analysis, the tubeworm backscatter signatures (dense bushes and semi-continuous bushes, Fig. 2.3) were distinguished as a separate group in the 2-group K-means solution. At that level (C-H statistic = 197.91) (not shown in the form of a table), most of the backscatter signatures of the dense (97%) and semi-continuous tubeworm bushes (86%) were grouped together. In the 4-group solution (C-H = 147.82) the clams and polychaete/limpet assemblages had about 40% of their signatures in a group that they shared with the broken sheet flow (Table 2.II). The clam and polychaete/limpet assemblages also had identical percentages (58%) in a second group (Table 2.II). The lava flow and the dense tubeworm bushes each had a major portion of their backscatter signatures in a single group (respectively 74% and 94%). The semi-continuous tubeworms habitat had a less distinct backscatter signature, with 25%, 61% and 13% of its echos divided among three groups. These preliminary results indicate that sonar backscatter analysis has considerable potential for mapping habitat distri-

bution in larger hydrothermal vent fields or seep environments where there is extensive coverage by tubeworm assemblages.

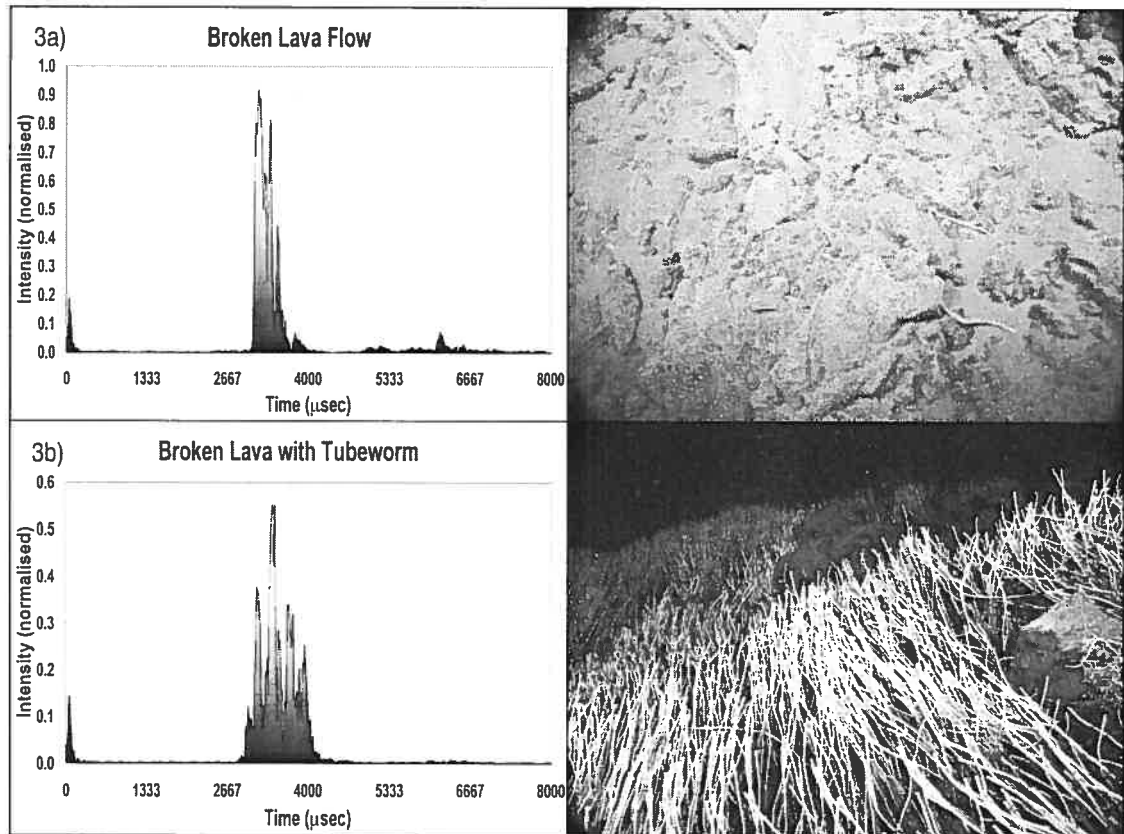


Figure 2.3: Example of contrasting habitat types and backscatter signatures at Clam Bed vent. a) Broken sheet lava flow signature (photograph of site on right). b) Acoustic signature of dense tubeworm bush (photograph on right). These sites were adjacent to each other on a ridge summit. The substratum beneath the tubeworm bush was similar to the broken sheet lava flow shown in (a).

Table 2.II: Results of K-means partitioning of backscatter signatures for the 5 studied habitats. The 2 and 4 group solutions represented local maxima of the Calinski-Harabasz criterion. The 4-group solution described in this table provided the best separation of habitat types. Data for each habitat type (columns of the table) indicate the percentage of the total number of backscatter signatures for that habitat that belong to each of the 4 groups. Values above 50% are in bold.

Grouping	Broken sheet flow	Dense tubeworms	Clams	Limpets	Semi-continuous tubeworms
1	15	3	58	58	13
2	74	0	42	40	1
3	8	3	0	2	61
4	4	94	0	0	25

At this point in our analyses, we do not know whether or not assemblages of smaller-vent organisms will be distinguishable from the surrounding basaltic substratum. Application of this method to more complex terrains will be more difficult. The Clam Bed site was selected for this study because it offered a relatively flat bottom and contains no mineral edifices. This avoided any confounding effects of relief on the backscatter signal.

The combination of mapping and GIS analysis allowed us to quantitatively compare the importance of the processes of *de novo* colonization and faunal disappearance at Cloud Vent in relation to shifts in assemblage composition or density. While we cannot exclude the possibility that intermediate stages occurred between our 1-year observation intervals, the data show that the rapid appearance and disappearance of fauna dominated the changes that occurred during the 4 years following the creation of Cloud Vent. In a study of another eruption and colonization event on the Juan de Fuca Ridge, Tunnicliffe et al. (1997) observed extensive blooms of bacterial mats followed by colonization by tubeworm assemblages. These authors suggested that microbial mats might be a necessary precursor to colonization of bare rock surfaces by tubeworms. In our 1998 video surveys, bacterial mats could be seen throughout the study area. However, the poor visibility caused by the voluminous venting of fluids rich in suspended

particles prevented the systematic mapping of bacterial mat distribution. By 1999, one year before the appearance of extensive tubeworm assemblages at Cloud Vent, the bacterial mats had largely disappeared. This observation would argue against the necessity of any precursor colonization by microbial mats before the arrival of tubeworms. However, the presence or absence of thin bacterial films that might attract tubeworm larvae cannot be confirmed from video observations.

## 2.4 Conclusion

Video mapping and sonar backscatter analysis are non-interventive techniques well suited to time series studies of the rapidly evolving faunal communities that develop around deep-sea hydrothermal vents. The integration of map data into a GIS, as briefly described here, greatly enhances the power of this information for the exploration of ecological questions. Sonar backscatter information, in addition to not being hampered by lighting and visibility problems, offers the advantage of permitting 3-dimensional reconstruction of habitat distribution on the seafloor using topographical and backscatter information. However, there are still a number of outstanding questions concerning the applicability of sonar backscatter studies to the hydrothermal vent environment, including habitat signature variability and the detection of patchy colonization of hard substrata that so often characterizes vent communities. We are presently investigating these questions at the Cloud Vent and Clam Bed sites.



## 2.5 Acknowledgements

This work would not have been possible without the field support of the NeMO Chief Scientist. R. Embley, the ROPOS submersible pilots and the crews of the NOAA Ship Ronald H. Brown, the CCGS John P. Tully, and the R/V Thomas G. Thompson. This research was sponsored by NSERC (Canada) Collaborative Research Opportunities grants to SKJ and PL, and by the Canadian Scientific Submersible Facility. ML was supported by an NSERC Canada Undergraduate Research Fellowship. We are particularly grateful to L. J. Hamilton (backscatter analysis), D. Marcotte (Kriging methods), M. Fellows (Imagenex surveys), J. Illman (navigation software training), and J. Marcus (faunal assemblage information).

---

## CHAPITRE 3

# SONAR BACKSCATTER DIFFERENTIATION OF DOMINANT MACROHABITAT TYPES IN A HYDROTHERMAL VENT FIELD

---

Sébastien DURAND<sup>1,2</sup> , Pierre LEGENDRE<sup>1</sup> and S. Kim JUNIPER<sup>2</sup>

Accepté pour publication dans "Ecological applications journal"

le 21 Août 2005 .

<sup>1</sup> *Département des Sciences Biologiques - Université de Montréal, C.P. 6128 succ.  
Centre-Ville-Montréal (Québec), H3C-3J7, Canada* <sup>2</sup> *GEOTOP - Université du  
Québec à Montréal, C.P. 8888 succ.*

*Centre-Ville-Montréal (Québec), H3C-3P8, Canada E-mail :*



### 3.1 Abstract

Over the past 20 years, sonar remote sensing has opened ways of acquiring new spatial information on seafloor habitat and ecosystem properties. While some researchers are presently working to improve sonar methods so that broad-scale high-definition surveys can be effectively conducted for management purpose, others are trying to use these surveying techniques in more local areas. Because ecosystem management is scale-dependent, there is a need to acquire spatio-temporal knowledge over various scales to bridge the gap between already-acquired point-source data and information available at broader scales.

Using a 675 kHz single-pencil-beam sonar mounted on the Remotely Operated Vehicle ROPOS, 2200 m deep on the Juan de Fuca Ridge, East Pacific Rise, five dominant habitat types located in a hydrothermal vent field were identified and characterized by their sonar signatures. The data, collected at different altitudes from 1 to 10 m above the seafloor, were depth-normalized. We compared three ways of handling the echoes embedded in the backscatters to detect and differentiate the five habitat types; we examined the influence of footprint size on the discrimination capacity of the three methods; and we identified key variables, derived from echoes, that characterize each habitat type. The first method used a set of variables describing echo shapes, the second method used as variables the power intensity values found within the echoes, whereas the last method combined all these variables.

Canonical discriminant analysis was used to discriminate among the five habitat types using the three methods. The discriminant models were constructed using 70% of the data while the remaining 30% were used for validation. The results showed that footprints 20 to 30 cm in diameter included a sufficient amount of spatial variation to make the sonar signatures sensitive to the habitat types, producing on average 82.1% correct classification. Smaller footprints produced lower percentages of correct classification; instead of the habitat types, the sonar data responded to intra-patch roughness and hardness characteristics. The sonar variables used in this study and the methods

for extracting and transforming them are fully described in this paper, and available in the public domain.

KEY WORDS : *canonical discriminant analysis; habitat types; hydrothermal vents; Juan de Fuca Ridge; Remotely Operated Vehicle (ROV); remote sensing; sonar backscatter.*

## 3.2 Introduction

Broad-scale remote sensing surveys have brought many benefits to agriculture, mineral exploration and environmental management in the terrestrial environment. In the field of landscape ecology, satellite imagery, airborne photography, hyper-spectral imagery, passive/active microwaves, radar, lidar systems, and so forth, provide information on habitat distribution, evolution, connectivity, structuring process, recovery rates, as well as evaluation of transition zones, meta-populations, and even plant population physiological status such as stress level (Carey et al., 2003; Hewitt et al., 2004; Kotchenova et al., 2004; Lee and Anagnostou, 2004; Lee and Chough, 2001; Moya et al., 2004; Schmidtlein and Sassin, 2004). However, because seawater is relatively opaque to electromagnetic waves (Foster-Smith and Sotheran, 2003), optical and radio-frequency remote sensing tools find little applications in ecological studies and habitat management on the deep ocean floor. Considering that more than 60% of the Earth's surface is covered by more than 1000 m of water, the lack of efficient investigation tools constitutes not only a serious obstacle to understanding the dynamics of biodiversity and ecosystem functioning at a global scale, but also impairs our capacity to correctly manage deep-sea resources. On the rare occasions where towed platforms and research submersibles reach such depths, mounted video and still cameras can be used to survey organism distribution and, at smaller scales, for direct estimates of benthic organism density, micro-topography and substrate characteristics. However, the spatial scope and resolving power of light-based systems remains very limited in what is essentially an aphotic and light-absorbing (i.e., turbid) environment. Consequently, detailed optical imaging of the deep seabed must be conducted at very slow speeds and can rarely be done from altitudes that would provide optimal spatial resolution (Parry et al., 2003, in situ experimentation). Benthic ecologists have long awaited the development of efficient remote sensing techniques that could be applied from a distance over large areas of the deep ocean floor (Hewitt et al., 2004). Even with the recent advances in light-based and laser-line systems (Carey et al., 2003; Irish and Lillycrop, 1999) developed for shallow

environments such as the coastal zone, acoustic technology remains the only method that has the potential to carry information through ultimately thousands of meters of water. Acoustic data also lend themselves much more easily to automated mapping and statistical analyses than do maps produced from underwater cameras.

Acoustic sounders and later-developed multi-beam sonars have been used extensively for precise, broad-scale mapping of seafloor topography. More recently, innovations in sonar technology have allowed researchers to demonstrate the potential for accurate mapping of seafloor habitat characteristics at broad scales. Through established methods (Burczynski, 2001; Burns et al., 1989; Chivers et al., 1990; Clarke and Hamilton, 1999; Ellingsen et al., 2002; Hamilton, 2001; Legendre et al., 2002; Prager et al., 1995), commercially available systems such as BioSonics' VBT™, Echoview™, QTC VIEW™, or RoxAnn™ are now used by researchers to extract habitat information from returning acoustic signals. One line of current research is the development of effective methods to cover large areas that would involve making the footprint of the sonar beam (sample area) large enough to reduce extrapolation needs. This would, however, result in loss of information on micro-scale habitat heterogeneity, which is of great importance for community ecology and for maintenance of biodiversity. Understanding ecological dynamics and managing ecosystems require the ability to effectively map broad expanses, as well as an understanding of smaller-scale ecosystem features, which quite often play a determining role at broader scales.

Using only acoustic spectral features, Pace and Gao (1988) successfully classified six seabed types: sand, mud, clay, gravel, stone, rock. Today, the most commonly used method is to use a side-scan sonar using frequencies around 1 to 200 kHz to detect the substratum type through the use of backscatter intensity curves and texture analysis of side-scan images (e.g., Brown et al., 2002; Zajac et al., 2003). But there has been little methodological development for backscatter interpretation at the higher acoustic frequencies that could distinguish seafloor habitats and macrofaunal communities.

This paper begins this methodological development by experimenting with acous-

tic returns from a high frequency sonar, in order to address the following questions: (1) Is the information found within backscatters informative enough to allow accurate discrimination of abyssal benthic habitats, in this particular case, five hydrothermal habitats within a vent field ecosystem? (2) If so, does increasing footprint size allow the acquisition of backscatters that are increasingly representative of the spatial heterogeneity inherent to each habitat? (3) Can we find specific sets of variables that could be used to correctly identify the nature of the habitat surveyed, based on sonar signatures?

### 3.3 Materials and methods

Acoustic information was obtained during dives of the remotely-operated vehicle (ROV) ROPOS in 2001 and 2002. Dives were conducted during cruises of the Canadian Coast Guard Ship J.P. Tully to the hydrothermal fields of the Endeavour Segment of the Juan de Fuca Ridge, in the northeast Pacific Ocean, 300 km southwest of Vancouver Island. Acoustic data were obtained using an Imagenex 881B single-beam sonar, equipped with a sub-miniature profiling head unit model 881-000-130, using 675 kHz frequency and a  $1.7^\circ$  pencil beam width, mounted on the submersible. Subsea positioning was determined using a long baseline acoustic navigation system (LBL) that included a PS8010 Edgetech transceiver and 5 bottom transponders georeferenced using the vessel's DGPS system. All interrogation, receiving and processing related for this LBL system was handled through the Seascope and Workboat software™ on the support vessel.

For all the local navigation and ground-truthing procedures, images from the submersible's low-light, Silicon Intensified Targeting (SIT) camera, as well as a 3-CCD color video were recorded on S-VHS or digital (Mini-DV) tapes for post-processing (mainly habitat identification and transect filtering).

To transform and statistically analyze the collected sonar information, we developed functions under the R-project version 2.1.0 (R Development Core Team 2005). R is a statistical language freely downloadable from the Internet at URL: <http://www.r-project.org/>.

#### 3.3.1 Study site and habitat description

Located 2200 m deep on the Endeavour Segment of Juan de Fuca Ridge ( $47^{\circ}57'47''\text{N}$  -  $129^{\circ}05'30''\text{W}$ ), Clambed is a hydrothermal vent field of circa 50 m by 20 m with a central, actively venting chimney (named "Hershey") standing 2-3 m tall and surrounded by localized diffuse venting. Covered mostly by broken lava flows and nearly sediment



free, the site's overall topography consists of two roughly parallel 2-3 m high north-south trending ridges colonized by hydrothermal vent tubeworms (*Ridgeia piscesae*) and polychaete/limpet assemblages (*Paralvinella palmiformis*, *Paralvinella sulfincola*) / (*Lepetodrilus fucensis*). Between the ridges, lightly sedimented depressions are found, which host small communities of vesicomid clams (*Calypptogena cf. pacifica*), another hydrothermal vent species. Based on previous sightings of the studied vent field and on work on Juan de Fuca ridge hydrothermal species assemblages by Sarrazin et al. (1999), five visually distinct and dominant habitats were selected to be probed *in situ*. In Fig. 3.1, a photograph of each habitat is associated with a sample sonar signature (first echo only).

1. Ridge top with dense, continuous tubeworm bushes (habitat code: Tube).— Within this key habitat, the structure complexity of the dense tubeworm communities has been hypothesized to be a leading factor in diversity (Tsurumi and Tunnicliffe, 2003). The channeling effect that tubeworm communities have on the hydrothermal fluid might reduce the environmental chemical and thermal fluctuations, just as the wind or temperature fluctuations are buffered in a terrestrial forest habitat. Within this microcosm, tubeworms can either serve as food source, refuge, or substratum, or as a hunting ground. Easy to distinguish from the other habitat types, these dense, white, bush-like structures are often covered by microbial mats (e.g., *Arcobacter sp.* and *Folliculina sp.*, (Léveillé and Juniper, 2003; Wirsen et al., 2002), and host many small worms and gastropods (e.g., *Lepidonotopodium piscesae*, *Paralvinella dela*, *Paralvinella palmiformis*, *Depressigyra globulus*), and squat lobsters (*Munidopsis alvisca*). Species residing within the tubeworm bushes are rarely visible in the video recordings.

2. Ridge top with semi-continuous tubeworm bushes (habitat code: Peri).— Found at the boundary of dense tubeworm bushes, this habitat exhibits low tube densities and is the most visually diverse of the five habitats selected. Species found both outside and inside the tubeworm colonies can be seen in recorded imagery. The substratum is clearly visible over about 50% of the area.

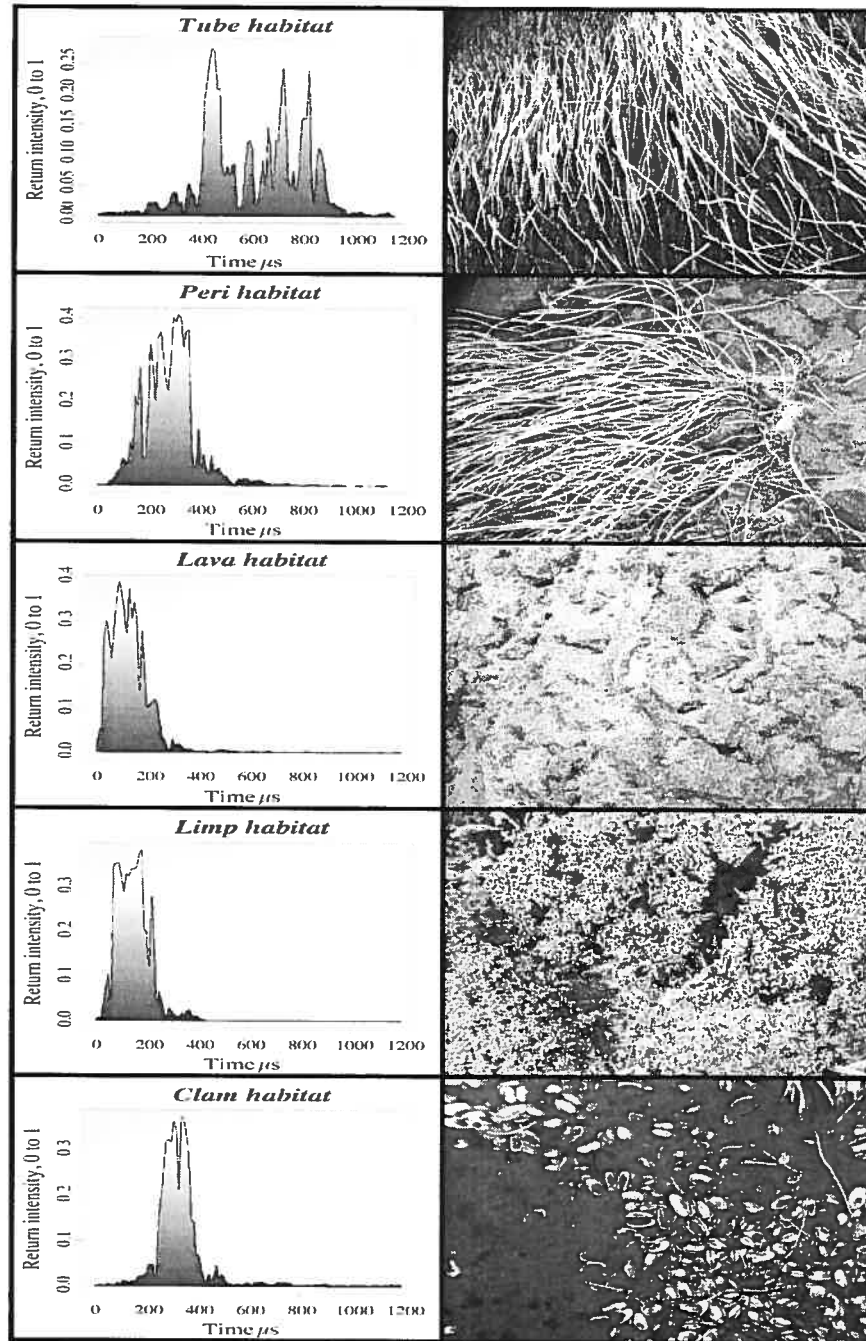


Figure 3.1: Sample echo signatures at an altitude of 8.5 m (left), corresponding to the habitat types pictured on the right. To represent the sample echo signatures, only the first echoes of the backscatters were used here; the first and second echoes are described in Fig. 3.2.

3. Ridge top without tubeworms (habitat code: Lava).—This habitat is the most common of all five habitats studied and it consists of bare to lightly-sedimented broken basaltic flow sheets. It is colonized by very low densities of species non-endemic to hydrothermal vents, such as unidentified holothurians, starfish, sponges, anemones, and a few crinoids.

4. Ridge top with polychaete/limpet assemblages (habitat code: Limp).—Usually located in the immediate vicinity of visibly intense hydrothermal flow emissions, these highly localized and dense communities of limpets and polychaetes can completely cover the underlying substratum. Some polychaetes (e.g., *Lepidonotopodium piscisae*, *Branchinotogluma grasslei*) can be seen attached to the small tubes and shells of *Paralvinella palmiformis*, *Paralvinella sulfincola* and *Lepetodrilus fucensis*.

5. Sedimented depression with clams (habitat code: Clam).—Within the most sedimented sections of Clambed, a mixture of clams, empty shells and a few fallen tubeworm tubes occur in sediment patches often visited by spider crabs (*Macroregonia macrochira*).

### 3.3.2 Study design

In a pilot study (Durand et al., 2002), we showed that sonar signatures were sensitive enough to differentiate geological and biological features based on their respective densities, textures and structures. To pursue the assessment of the use of sonar signatures as a remote sensing tool requires some basic assumptions to be made: (1) Local water variations in chemistry, temperature, and amount of suspended particles did not significantly influence backscatters. (2) ROV submarine noises, gear, or transponder sonar installations did not interfere with the sonar acquisition process. (3) For every acquired backscatter, the area of seafloor receiving the sonar interrogation ping was horizontal and flat (Clarke and Hamilton, 1999).

To assess the discrimination power of our sonar system, sonar data were collected at different altitudes from 1 to 10 m above the seafloor, using vertical transects. At

least three vertical transects representing each habitat were taken at different locations in the field. With the sonar head always pointing straight down (at zero degree angle), each vertical transect started in a stable and controlled position with the ROV resting on top of the selected site. After a short acquisition period, henceforth referred to as the stable section, a slow and controlled vertical ascent was instigated. During sonar acquisition, both the SIT and color video cameras were recording. Once beyond 5 m of altitude, the pilot increased the submarine ascent speed to minimize unwanted horizontal drifting. Drifting is caused in part by the increased speed of horizontal currents as the submersible rises above the seafloor; it is also a direct consequence of visibility deterioration with altitude, which interferes with piloting. Using reference points such as key geological structures and other objects visually identified at the beginning of each transect recording, the transect positioning and good visibility could be ensured only until an altitude of 10 m. None of the data acquired beyond this altitude limit were used since the positioning was too uncertain and prone to error. A total of 18 vertical transects were acquired, with an average duration of 4.5 min and about 670 sonar pings per transect. In 52 min 53 sec of raw recording, 7761 backscatters were recorded in digital form with a mean sampling interval of 0.41 sec, within the 1 to 10 m altitude range.

### **3.3.3 Signal processing methods**

#### *Visual filtering and data treatments*

With the footprints localized on SIT video frames, each sonar transect was visually filtered. Transect segments where the footprints were outside the targeted habitat and frames with poor visibility were eliminated. Using an algorithm developed for this research and described in the next two subsections, the raw backscatters of the retained transect portions were analyzed. The beginning of the first echo was located by scanning the backscatters for their first substantial intensity increase. These areas

were then used to estimate the sonar head acquisition altitude during sampling; Fig. 3.2 describes how the backscatters were segmented.

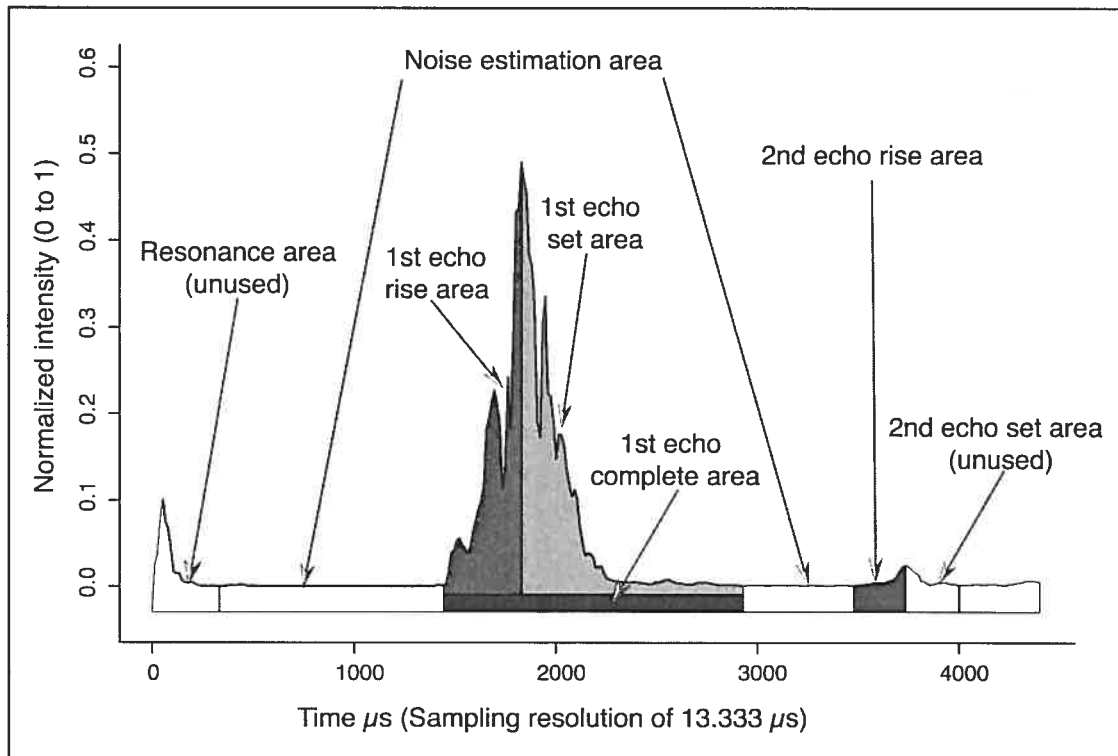


Figure 3.2: Echo segmentation of a real backscatter, acquired at 1.28 m of altitude, for which intensities have only been filtered for noise. The intensity variables described in Fig. 3.3 come from the 1<sup>st</sup> echo complete area. The other variables were extracted using both the first and second echoes.

Using these altitude estimates, a filtering procedure was initiated to identify and remove any other obviously bad signals carrying incomplete echoes or erroneous intensity curves. To detect and locate the first and second echoes (Fig. 3.2) inside a backscatter, both the original and smoothed backscatter were used; the smoothing algorithm used a moving window averaging 21 consecutive intensity values. We removed any acoustic return for which detection of either the first or second echo failed. For the remaining 7646 acoustic returns, we subtracted from the whole echo the signal ambient noise, averaged from the noise estimation areas. For logistic reasons, the recording of a large number of the echoes stopped within the second echo set area. As a result, derived

variables that express proportions between the first and second echoes consider only the second echo rise area.

Since our data were acquired at different altitudes, to compare echoes with one another, a depth normalization procedure was applied to correct both the strength of the recorded intensity values and the temporal spreading of the backscatters (Clarke and Hamilton, 1999; Hamilton, 2001). Note that since the sonar beam is a cone, the size of the sampled area, or footprint, is physically linked to the acquisition altitude. Therefore, even with depth normalization that accounts for temporal spreading and power, it is not possible to compensate for the effect of insonifying a large versus a small habitat area. To assess the impact of such variation, altitude-dependent data tables, described below, were extracted and analyzed.

### *Sonar variable extraction methods*

Three approaches were used to extract variables from backscatters. The first approach is similar to the methods used in the QTC VIEW™, RoxAnn™, and BioSonics™ software. We computed a series of variables from both the first and second echo sections of all depth-normalized backscatters, producing a data table of 28 extracted variables referred to as VE (not shown), describing locations, sections, or proportions of areas. These variables are described in Appendix A, Table A.I.

The second method used only the intensities of the first echo as variables. Once depth normalization was applied, the associated temporal correction stretched or compressed the individual acoustic returns. Intensities had originally been recorded at 13.333  $\mu$ s intervals; the intensities making up the first echo were resampled with a 10  $\mu$ s sampling interval, after smoothing using the “interspline” function of the R language (R::package::splines::interspline). Using these equally-spaced intensity values, we created our intensity variables (Int) by averaging sets of five consecutive values, and we assembled them in a data table referred to as VI (not shown), which contained a maximum of 92 consecutive intensity variables per backscatter. For shorter echoes not

producing 92 variables, zeros filled the empty cells. The third method combined all the variables found in VE and VI into a new data table called VT.

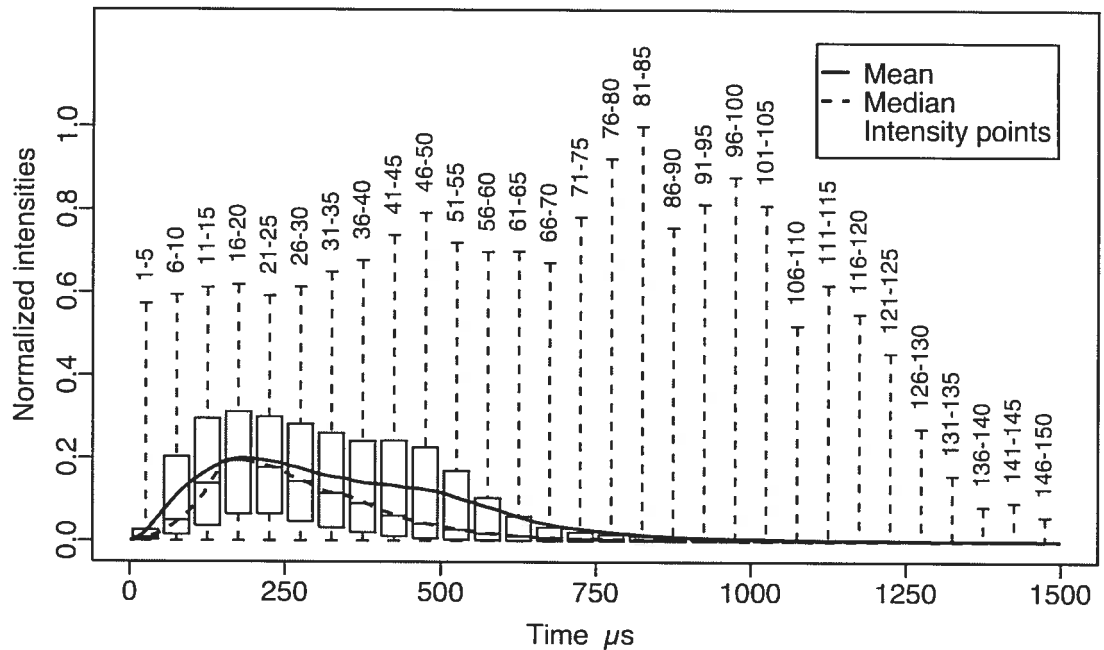


Figure 3.3: Picture combining the intensity values of all backscatters used in this study. The cloud of grey stars represents the first 150 intensity values found in the first echo of all acoustic returns. Each box plot portrays the localized distribution of five consecutive intensity values, which form one intensity variable. The whiskers of the box plots show the minimum and maximum values, while the boxes show subsection quartiles (respectively, 25, 50 and 75%).

In Fig. 3.3, each box plot gives a descriptive insight into the role of the first 30 intensity variables used in the VI and VT data tables. By combining the box plot positions and the mean and median lines, it is possible to visually characterize the variables. The variables averaged from the corresponding intensity groups 1-5, 6-10 and 11-15, called “Int1-5”, “Int6-10” and “Int11-15” respectively, clearly represent the rising portion of the echoes. Variables “Int16-20” and “Int21-25” describe the first echo peak. The intensity values from 26 to 90 (intensity variables “Int26-30” to “Int86-90”) correspond to the backscatter tail or set area of the first echo. Beyond that, the curve becomes flat and cannot be visually interpreted.

### *Data filtering, segmentation and transformations*

In data tables VI, VE and VT, any variable exhibiting no variation within either of our habitat types, as well as any variable showing very little variation, or less than 10 individual values differing from the mean, was removed. The first 59 variables were kept in data table VI (“Int1-5” to “Int291-295”); 25 variables remained in data table VE after “Histo5” to “Histo7” had been removed (Table A.I).

In order to bring the data tables close to the multinormality condition, which would improve the performance of discriminant analysis (next section), different transformations were applied to the VI and VE data tables prior to the creation of table VT. In data table VI, 15 possible transformations described in Table A.II of Appendix A were tried in turn on six different subtables. Because all variables found in data table VI are of the same nature (they represent signal intensities), they should all be subjected to the same transformation. To estimate the common skewness of all transects, we standardized the values of each variable within each transect, which controls for the effect of the first two moments of their distributions, and combined the standardized values in a single table. The absolute values of skewness were averaged across variables for each VI data table. The transformation that produced the smallest mean skewness was selected.

For data table VE, all variables were not of the same nature and did not have similar distributions. Therefore, for each variable we tested the following: no transformation, the square root, the double square root or the log transformation, and selected the transformation that produced the smallest skewness. After applying the best transformation to each variable, all variables found in the VI and VE sets were combined to create VT data tables.

For each of the resulting and newly transformed VI, VE and VT data tables, 5 new subtables were created. Using only the backscatters found in the stable section (code STB, about 1 m altitude) of each transect, STB subtables were constructed for each habitat and used for control and initial tests. This application was restricted to the



STB subtables because all their backscatters had been acquired during the period of greatest visibility, ensuring accurate habitat identification; altitude variations were also minimum so that the acquisition of acoustic returns was unaffected by depth-related phenomena (see Clarke and Hamilton, 1999). Then, three altitude-related subtables were created using the following backscatter altitude acquisition ranges: 1 to 4 m, 4 to 7 m, and 7 to 10 m. Finally, for each of the five habitats, the backscatters found in the best altitude transect, in terms of visual sample quality and ROV displacement, were used to produce the BEST subtables. We then assessed the discriminating power of our three sets of variables found in the VI, VE and VT data tables, and the effect of altitude and footprint size, by combining the information provided by the analysis of all these subtables.

### *Statistical analyses*

To assess the discriminating power of the sonar variables found in the VI, VE, and VT data tables, the percentage of correct classification (PCC) after discriminant analysis was computed using the function `R-Pkg::MASS::predict.lda`. For each data table, a random selection representing 70% of the backscatters was used to compute the discriminant model while the remaining 30% served to predict the habitat associated with each backscatter. The PCC index was calculated for these validation data.

For each data table, variance condensation was achieved by principal component analysis (PCA). We used only the principal components accounting for 99% of the variance (assembled in the COM data sets) in linear discriminant analyses (`R-Pkg::MASS::lda`) and obtained our first PCC results without selection of wave form variables; see Table 3.I.

Table 3.I: Percentages of correct classification.

Variable Set	Methods	Tables and subtables						Mean
		STB	BEST	ALL	1-4	4-7	7-10	
VI	COM	85.4	69.9	58.7	63.1	63.0	72.4	68.8
	FWD	79.6	71.0	58.1	61.3	61.7	73.1	67.5
	SEL	83.8	71.6	58.6	62.8	63.3	76.9	69.5
VE	COM	94.6	78.2	65.3	67.9	67.7	77.6	75.2
	FWD	90.8	72.9	61.0	65.7	59.0	68.7	69.7
	SEL	86.2	74.7	58.5	63.5	62.4	68.7	69.0
VT	COM	95.8	78.8	68.0	70.7	76.6	85.1	79.2
	FWD	94.4	76.9	63.4	68.7	65.5	76.9	74.3
	SEL	94.1	75.5	63.6	70.4	68.4	84.3	76.1
VI mean		82.9	70.8	58.5	62.4	62.7	74.1	68.6
VE mean		90.5	75.3	61.6	65.7	63.0	71.7	71.3
VT mean		94.8	77.1	65.0	69.9	70.2	82.1	76.5
Total mean		89.4	74.4	61.7	66.0	65.3	76.0	72.1

*Note:* Percentages of correct classifications (PCC) for the variable sets found in the VI, VE and VT data tables, were estimated by three methods. The first analysis (COM) used the complete set of principal components accounting for 99% of the variance in the data. The second analysis (FWD) used only the original variables retained by forward selection. Finally, analysis SEL used 8 VI, 8 VE, or all of these 16 variables, depending on which variable set was tested. This variable selection was based on the distribution and frequency of the variables previously selected by forward selection. Analyses were repeated using the acoustic returns obtained when the ROV was stable at low altitude (STB), the transects with the best backscatter quality (BEST), all backscatters in all vertical transects (ALL), or only the backscatters from altitudes of 1 to 4 m (1-4), 4 to 7 m (4-7), or 7 to 10 m (7-10).

Alternative strategies were also used. Firstly, instead of computing a PCA for each VI, VE, and VT data table, the variables with the highest contributions were identified by forward selection with permutation tests. A function to carry out discriminant analysis, following the algorithm described by ter Braak and Smilauer, 2002, section 3.11, with forward selection of explanatory variables (*ibid.*, section 5.8.1), was developed in the R language by S. Dray (CNRS, Lyon, France). Using only this selection of variables, henceforth referred to as FWD, discriminant analysis was computed again, producing another set of percentages of correct classifications describing the discrimi-

nating power of a smaller set of selected variables for each of the three kinds of data tables (VI, VE, and VT). Secondly, since the set of selected variables varied from table to table, identical sets of variables had to be used in all tables to allow comparisons and understand the role of key variables. By choosing the variables with the highest selection frequencies in all FWD selections (Appendix A, Table A.III), we created a subset of three variables called SEL. Discriminant analyses were computed with it, and a series of explanatory discriminant analysis plots were produced. (Appendix B).

## 3.4 Results

Depth variation has been shown in the literature to affect our capacity to detect and differentiate sonar signatures (Hewitt et al., 2004). Without a proper depth normalization procedure, the effects of uncorrected altitude fluctuations are likely to overshadow the variation inherent in the nature of the seabed, and fatally link the sonar signatures to altitude-related variables. To perform an accurate depth normalization, since the rate at which the sound is absorbed as it travels through water was estimated instead of being precisely measured, corrective measures were taken.

To adjust the absorption rate and consequently optimize our depth normalization procedure, we used several plots like those presented in Fig. 3.4 to visualize the effect of the power correction by comparing original to the power-normalized backscatters, respectively drawn in Figs. 3.4a and 3.4b. The thick and dark grey lines shown in both plots represent strong intensities and correspond to the first and second echo areas. In Fig 3.4a, as the ROPOS gained altitude, the intensity of the backscatters weakened from left to right in all transects; consequently, paler grays are showing in the right-hand portion of each transect. In Fig. 3.4b, the power normalization algorithm removed the fading, to a point where a homogeneous grey background was found, from left to right, within and among transects. The darker curves look more homogeneous; this is a visual sign of an accurate power correction (meaning as in Hamilton, 2001). Comparison of Figs. 3.4a and 3.4b shows that the power correction in the depth normalization procedure increased our capacity to visualize specific sections of the backscatters. (In Fig. 3.4b, the addition of three overlaid white lines to the intensity profiles describing the first echo start and end positions plus the second echo start positions served two purposes. First, they allowed us to visually assess the success of the echo detection algorithm; secondly, they served as visual markers showing from which sections of the backscatters we are extracting our variables.)

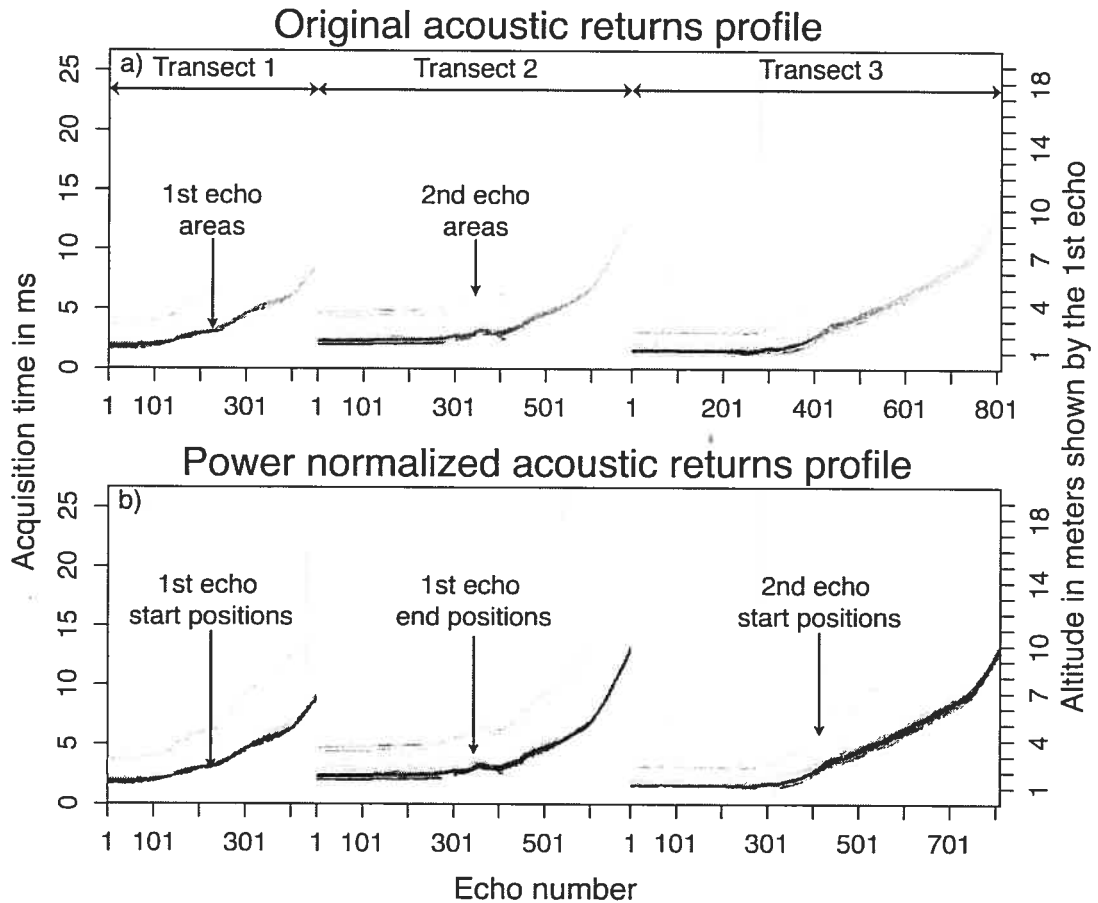


Figure 3.4: a) Original profiles of recorded acoustic returns. Echoes are represented by vertical lines of pixels going from the bottom to the top of the graph and shaded according to signal intensity (darker is stronger signal), for three vertical transects in the Peri habitat. The three dark curves are areas of strong intensities and correspond to the peaks of the first echoes. Above these three curves are three paler grey curves corresponding to the peaks of the second echoes. The rising shape of these lines, from left to right along each transect, is caused by the altitude gain. When increasing the altitude, the delay between the time when the signal is sent and received for the first time by the sonar head also increases. Consequently, the resulting sonar signal intensity power is weaker because the longer it travels in water, the more it gets dissipated and absorbed. b) The power-normalized acoustic returns are now uniform in shading. White lines show the boundaries of the first echoes and the second echo start positions, detected by the algorithm. The grey lines of the second echoes are mostly hidden by the white lines drawn over them. The second echo end positions are, in our case, the second echo maxima. The second echo end positions were not shown; they would have been barely distinguishable because they were too close to the second echo start positions.

The first and second echo starting curves should, in principle, be smooth if the algorithm operates correctly, because the physical conditions under which the echoes were acquired involved slow and gradual ROV rise and constant recording. This is the case for the first echoes, but the detection of the beginning of the second echoes is more random. Instead of having the second echo bouncing off the seafloor to the air/water interface to the seafloor and finally to the sonar head, our second echoes are reflected on the seafloor, the underside of the ROV, and the seafloor again, before reaching the sonar head for the second time. Reflections from a large, homogeneous air/water interface are much smoother than reflections from the ventral surface of the ROV, which has an irregular shape, holes, and attached equipment such as canisters of various shapes, textures and densities. In addition, because the second echoes are by nature weak, an accurate second echo detection algorithm was difficult to produce. More work will be necessary to improve this algorithm.

After extraction of the variables, normalizing transformations were applied to all data tables. Table A.I shows the VE variable names and gives the selected transformations and associated skewness values. For the VI intensity variables, Table A.II gives the skewness values calculated for all transformations applied to segments of the altitude transects. Exhibiting the lowest skewness values, the double square root followed by the arcsine transformation was consistently the most appropriate combination of transformations for VI variables, except for the 7-10 subtable where the Hellinger transformation produced a result slightly better than the arcsine transformation. A double square root followed by an arcsine transformation was consequently applied to all VI data tables (intensities). That transformation implies that only the shape information remains to be analyzed in the transformed VI data tables, since VI variables are ranged by dividing their values by the maximum intensity present in the original signal.

Examining the VI, VE, and VT variables retained after forward selection, the following trends were observed (Table A.III, Appendix A). On average, out of 59 VI, 25 VE,

and 84 VT available variables, only 6, 7, and 9 variables, respectively, were retained by forward selection. The frequency distribution of the intensity variables selected in either VI or VT shows that 20% are found between Int1 and Int26, 60% are found between Int26 and Int90 and 20% are found above Int90 (the latter group was never selected at low altitude, 1-4 m). Consequently inside the group of the most often selected variables (code SEL), a similar ratio was kept: the 8 Int variables selected were “Int11-15”, “Int21-25”, “Int31-35”, “Int46-50”, “Int55-60”, “Int71-75”, “Int91-95”, and “Int181-185”; for the VE variables retained after VE or VT forward selection, the 7 variables with the highest selection frequencies were “DRSx”, “Skew”, “NewAlt”, “Vmn.s”, “Histo1”, “Vmx.sE1”, “Vmx.sE2”, “Time.RE1”.

Table 3.I shows differences in classification performance among the three types of data tables. Data tables VT lead to higher percentages of correct classification than either VI or VE. Table 3.I allows us to answer our first question: Is the information found within backscatters informative enough to allow accurate discrimination of abyssal benthic habitats, in this particular case, five hydrothermal habitats within a vent field ecosystem? Using only the sonar samples taken during the stable section of each transect, defined by moments of low altitude where the ROV was minimizing its vertical and horizontal displacements, a control test was performed. Even if the overall efficiency of our method cannot be assessed using only the percentages of correct classification for STB data, the high PCC values (82.9, 90.5 and 94.8%) obtained in this controlled situation confirm that the five hydrothermal habitats under study possess differentiable and repeatable sonar signatures.

### 3.4.1 Relationship between PCC and altitude

To verify the performance of the VI, VE, and VT data tables between 1 and 10 m, we compared the percentages of correct classification for all backscatters (code ALL) to those of the transect with the best backscatter quality (code BEST). The latter gave, on average, 12.7% better results (Table 3.I). It appears that this subset of variables

displayed limited habitat variation in the sonar signatures and consequently facilitated discrimination. In order to understand why, on average, ALL and BEST had such low discrimination capacity (respectively 61.7 and 74.4%) compared with the mean of 89.4% for the low-altitude subtable (code STB), we compared the PCC obtained for altitude-specific subtables (1-4, 4-7, and 7-10 m) to identify how the intra-habitat variation was distributed within our transects. At altitude ranges of 1 to 4 and 4 to 7 m, on average, the PCC's obtained were similar, but an increase in PCC averaging 10.7% occurred between 4-7 and 7-10 m.

In an attempt to illustrate why an altitude-related variation can be seen even on depth-normalized data, we compared the classification obtained through sonar signature discrimination with our initial visual habitat classification for the VT data tables. The results in Table 3.II provide an answer to our second question: Does increasing footprint size allow the acquisition of echoes that are increasingly representative of the spatial heterogeneity inherent to each habitat? At 1 to 4 m, 4 to 7 m, and 7 to 10 m of altitude, the sonar beam width was respectively 3-12, 12-20, and 20-30 cm in diameter. This means that the very small footprints at low altitude were more likely to detect intra-habitat patches of different textures and densities. For example, the classification obtained for the Peri habitat at 1 to 4 m shows that, besides the 68.4% of the backscatters that were correctly classified, most of the remaining acoustic returns were classified as representing the Lava and Tube habitats, which are Peri main constituents. In the classification results obtained for the 1 to 4 m and the 4 to 7 m data, a clear division exists between the backscatters belonging to the Lava, Peri, and Tube habitats on the one hand, and the Limp and Clam habitats on the other. The 4 to 7 m data do better than the 1 to 4 m data at separating the Limp and Clam backscatters. At 7 to 10 m of altitude, most (91.8%) of the sonar signatures were correctly classified, indicating that an optimal footprint size had been reached.



Table 3.II: Habitat assignments of all echoes based on the VT data tables, using the complete variable sets (COM).

Altitude range	Habitat observed	Habitat assigned by sonar					Partial PCC
		Tube	Peri	Lava	Limp	Clam	
1-4	Tube	859	228	30	4	2	76.5
	Peri	184	996	166	48	63	68.4
	Lava	3	127	780	28	48	79.1
	Limp	0	8	15	806	250	74.7
	Clam	0	42	27	220	772	72.8
4-7	Tube	373	55	14	0	3	83.8
	Peri	43	271	68	4	6	69.1
	Lava	28	61	176	2	3	65.2
	Limp	0	4	1	231	5	95.9
	Clam	1	9	9	13	119	78.8
7-10	Tube	113	2	0	0	0	98.3
	Peri	2	103	0	1	0	97.2
	Lava	3	0	71	11	0	83.5
	Limp	0	2	15	89	0	84.0
	Clam	0	0	0	0	29	100.0

*Note:* To construct this table, all available backscatters and variables of the VT data table were used to create the discriminant model and for prediction. The total number of backscatters in the data tables are: 5706 for 1-4 m, 1499 for 4-7 m, and 441 for 7-10 m. Partial PCC stands for the percentage of backscatters from a given habitat that were correctly classified by reference to our visual habitat classification.

## 3.5 Technical discussion

### 3.5.1 The influence of altitude

Having shown (Fig. 3.4) that an accurate depth normalization was applied on all backscatters, because the footprint size of the sonar beam is directly related to the ROV altitude by physical laws, the variable “NewAlt” describing the ROV altitude was used to monitor the impact of footprint size on our discrimination capacity. By looking at the explanatory variables selected to describe the data tables and subtables (Table A.III), we realized that in the 7-10 subtables, “NewAlt” was never selected among the significant variables for VE and VT. This absence was attributed to the fact that, as the footprint expands with altitude, the backscatters signals incorporating more and more of the habitat’s fine-grain spatial heterogeneity. Therefore, as long as the amount of heterogeneity sampled is not sufficiently representative of a sampled habitat texture and density, the nature of the information in the backscatter is likely to change with altitude. Thus, as long as the sonar has sampled an area representative of the habitat general texture and density, whatever the variation in altitude, the resulting sonar signature variation no longer relates to altitude, producing better discrimination among habitat types.

### 3.5.2 The nature of the five habitat type signatures

Our third and last question was: Can we find specific sets of variables that could be used to correctly identify the nature of the habitat surveyed, based on sonar signatures? We used the set of selected variables (SEL) to compute discriminant functions among habitat types and produced six graphical representations of the resulting habitat cluster projections on the first and second discriminant axes (Appendix B). From these analyses, the centroids of the habitat clusters were correlated with the environmental variables. The correlations were noted in Table 3.III as either positive “+”, negative “-”, or null “0”.

Table 3.III: Correlation between sonar variables and habitat types, for the three altitude ranges.

VE variable	Habitat types						VI variable	Habitat types					
	Alt	Tube	Peri	Lava	Limp	Clam		Alt	Tube	Peri	Lava	Limp	Clam
DRSx	1-4	-	-	0	+	+	1-4	-	-	+	+	+	
	4-7	-	0	0	+	+	4-7	-	0	-	+	0	
	7-10	-	-	+	+	+	7-10	0	-	+	0	0	
Histo1	1-4	+	+	+	-	-	1-4	0	-	+	0	0	
	4-7	+	-	-	-	+	4-7	+	0	+	-	+	
	7-10	0	0	0	0	0	7-10	0	+	-	-	+	
Skew	1-4	-	-	0	+	+	1-4	+	0	+	-	-	
	4-7	-	-	-	+	+	4-7	+	+	+	-	-	
	7-10	-	-	+	0	+	7-10	0	+	-	0	0	
Time.RE1	1-4	+	0	-	0	-	1-4	+	+	-	-	-	
	4-7	+	0	0	-	-	4-7	+	+	0	-	-	
	7-10	+	+	0	0	-	7-10	+	+	-	-	0	
Vmn.s	1-4	-	0	+	-	0	1-4	+	+	+	-	-	
	4-7	+	-	-	-	+	4-7	+	+	+	-	-	
	7-10	0	-	0	+	+	7-10	+	0	0	-	-	
Vmx.sE1	1-4	-	+	+	-	-	1-4	+	0	+	-	-	
	4-7	0	+	+	0	-	4-7	+	+	+	-	-	
	7-10	0	0	0	0	0	7-10	+	0	0	0	0	
1-4	-	0	+	-	0	1-4	+	0	+	-	-		

Continued on next page

Table 3.III – continued from previous page

VE variable	Habitat types					VI variable	Habitat types					
	Alt	Tube	Peri	Lava	Limp		Clam	Alt	Tube	Peri	Lava	Limp
Vmx.sE2	4-7	-	+	+	+	-	4-7	0	0	0	0	+
	7-10	0	0	0	+	0	7-10	+	0	0	-	-
	1-4	0	0	0	0	0	1-4	0	0	+	0	0
NewAlt	4-7	+	+	+	0	-	4-7	0	-	0	-	+
	7-10	0	0	0	0	0	7-10	-	+	-	0	+

*Note:* Correlations of selected environmental variables with the centroids of the five habitat types in the space of the first two discriminant functions, for the various altitude subtables 1-4, 4-7 and 7-10 m. Positive, negative, and uncertain correlations are respectively marked as “+”, “-”, and “0”. This latter case occurred when the habitat centroid was at an angle close to 90° with the variable vector or when the variable was considered unstable based on its canonical weight and correlation vectors shown in panels a and b of Figures B.1 to B.6, Appendix B. For each habitat, the variables showing the same correlation sign over the three altitude classes are highlighted in dark grey; those that only have 2 identical signs out of 3 are highlighted in light grey.

The VE and VI results presented in Table 3.III were written to two data files, each with 5 rows (habitat types) and 24 columns (the rows of Table 3.III), and analyzed by *K*-means partitioning. For VE and VI as well, the results indicated the presence of two major groups of habitats differentiated by the variables derived from the backscatters: Clam and Limp formed the first group, Lava, Peri and Tube formed the second. Skewness of the first echo (variable "Skew"), as well as intensity variables "Int31-35", "Int46-50", and "Int71-75", were good indicators of this partition.

### *Clam and Limp habitats*

Habitat Clam had the most highly and positively skewed first echo, followed by Limp. The rise section of the first echo was short and correlated with strong intensities whereas the set section had mostly low intensities. Clam's sonar signatures were also negatively correlated to the maximum value of the smoothed first echo (variable "Vmx.sE1"), which indicates the presence of a smooth and soft type of surface (Bax et al., 1999). The positive correlations of the Clam centroid with the minimum value between the two echoes (variable "Vmn.s") indicates that after the first echo, the ambient noise remaining in the signal was higher than for other habitats.

The less strongly skewed signatures of the Limp habitat showed a small amount of low-class intensities (variable "Histo1"), very strong negative correlations with the minimum value between the two echoes (variable "Vmn.s"), and good positive correlations with the maximum value of the smoothed second echo (variable "Vmx.sE2"). These correlations support the idea that the Limp habitat contained high densities of gastropod shells, which produced a very reflective, hard and smooth surface allowing for low energy penetration. The sonar wave dissipated well, which produced low intensity values between the two echoes (variable "Vmn.s"). An interesting fact about Limp is the differentiation between the rise and set sections of the first echo. As soon as the positively correlated rise section passed the "Int11-15" intensity variable, negative correlations appeared from that point until the end of the set section (variable "Int91-95").

### *Tube, Peri and Lava habitats*

In the group with the most negatively skewed first echoes, the Tube habitat was the most extreme, followed by Peri sonar signatures. In both case, the echo shapes were the opposite of Clam and Limp. Their intensity variable correlations described a slow rise section, followed by higher intensities in the set section. While Peri signature presented positive correlations in the set section of the first echo until variable "Int56-60", the Tube correlations were positive until "Int91-95"; they were more stable in the sense that they were found more often in the entire 1 to 10 m altitude range. This suggest that, as the density of tubeworms increased, the first echo became longer since the positive correlations with intensity variables went further to the right in the set section. Beside the fact that the Tube's first echo intensities could be described by low intensities, they were shaped by numerous peaks and troughs. The positive correlations with variable "Histo1" and negative correlations with "Vmx.sE2" indicate how weak the sonar signal became after multiple reflections around the uneven and smooth tubular structures of the tubeworms.

Lava represented an intermediate case between the Tube and Clam extremes. In terms of correlations, Peri sonar signatures were intermediates between the Lava and Tube signatures. Obviously affected by the presence of the relatively rough and hard lava surface within its habitat, most of the strong correlations seen in Tube, such as with "Skew", "DRSx", "Histo1", or in Lava with "Vmx.sE1", were weaker in the Peri habitat. The Lava sonar signatures correlated with a quick rise and a set section that showed positive correlations reaching up to "Int71-75". It was also the habitat associated with the strongest smoothed maximum in the first echo. The high-intensity set section of Lava might relate to the fact that most Lava reflections were influenced by the unevenness of the broken lava sheets.

## 3.6 Conclusion

Beside surface roughness and hardness, many factors such as sonar signal frequency, ping length and beam width (footprint) can affect echo shapes. One of the major issues in backscatter analysis is the use of correct depth normalization procedures (Hamilton, 2001). To properly study the reflective nature of each habitat, we must ensure that most of the altitude-related variation is removed prior to analysis. The visual representation of that correction, such as in our Fig. 3.4, is quite important because it allows an assessment of the procedure used.

Under the assumption of an accurate depth normalization procedure, the presence of the “NewAlt” variable among the variables retained by forward selection would indicate the influence of footprint size variation. “NewAlt” was not selected to describe any of the 7 to 10 m data tables (Table A.III, Appendix A). That, and the stronger discrimination shown by the 7-10 m tables compared to the other depths (Table 3.II), led us to conclude that footprint size can drastically affect our capacity to investigate and ultimately detect habitat types.

Prior to any sonar survey, it is essential to make sure that the sonar settings are optimal. We used vertical transects over identifiable habitat types to verify the sonar’s ability to differentiate the five habitats under investigation. This exercise permitted the identification of key variables derived from backscatters and allowed us to identify an optimal footprint size to achieve sampling at scales that are representative of the general habitat textures and densities. Having optimized the sonar acquisition settings and depth normalization procedure, to bring even more robustness to sonar surveys, new sonar technology will need to be developed to allow the footprint size to remain constant during seafloor classification surveys (Legendre et al., 2002). Even when that technology becomes available, we will still be a long way from developing databases of sonar signature definitions describing diverse habitat types found over the whole benthic ecosystems. To construct such a database would require each habitat to be described using a constant set of variables based on various and specific sets of frequencies, foot-

print sizes, and pulse lengths. Before such standardization can be initiated, more work will be required to identify the best frequency combinations and sets of explanatory variables.

In this paper, we have shown that abyssal habitat identification is possible through the use of sonar signatures based on only one frequency, using a small and changing footprint, and using a remotely operated vehicle operating (ROV) at 2200 m deep. This provides some optimism for future sonar mapping developments. Multi-scale high-resolution seafloor sonar surveys may prove very useful for habitat mapping, resource evaluation and ecosystem management purposes.

Before sonar-based systems can be used as routinely for broad-scale surveys of habitats, many issues remain to be addressed both in terms of the sonar signal frequencies to be used and the establishment of key variable sets. Bax et al., 1999 wrote: "...it is clear that the full power of acoustic habitat discrimination has not yet been realized – there is far more information in the returning echoes and the pattern of echoes than is currently being interpreted."

Sonar remote sensing surveys require both a set of sonar signatures and some ground-truthing, the latter through either visual investigation or physical sampling, both of which are highly time consuming. The need for ground truthing could be reduced through the development of a database on the behavior of sonar signatures in various types of substrata and habitats through a series of criteria spreading over ranges of specific frequencies and sampling unit sizes (grain size). The development of such a database would require cooperation between researchers and the companies providing benthic remote sensing services. In order to encourage free and open communication and debate, we provide the definitions of our variables in Appendix A for scrutiny and use by the scientific and technology communities.

The sonar variables developed in this study and the method for extracting and transforming them are fully described in this paper and are available in the public domain.



### 3.6.1 Technical implications for other domains

Classical remote sensing methods are extensively used to produce bathymetric maps describing the demersal relief found in any type of water body. In these surveys, variables describing the echo time of arrival, like “NewAlt”, are used to compute altitude, which is, when added to the sonar depth, the information illustrated in bathymetric maps. With variables such as the echo general power intensity, e.g. “Vmx.sE1”, texture and density layers can be overlaid over bathymetric maps for substrate type identification. Extending the domain of application further, the method presented in this paper can be used as a guide for those who either wish to extract more information out of remote sonar surveys, find other useful variables to extract, or use new sonar frequencies. The science behind understanding sonar signatures is young, but it has potential applications in fine to broad-scale ecological surveys serving monitoring, management, and exploration purposes. Fish school identification capabilities could be improved by using some of the sonar variables described in this paper. Beyond the realm of aquatic sciences, sonar signatures can be used in many terrestrial applications. Mobile robots, which are already extensively using ultrasounds, have external sensors; a fine analysis of the sound returns in detection algorithm would give robots another mean to identify the nature of the objects they encounter.

## 3.7 Acknowledgments

This work would not have been possible without the dedication of the ROPOS submersible pilots and the crewmembers of the CCGS John P. Tully. This research was sponsored by NSERC (Canada) Collaborative Research Opportunities grants to S. K. Juniper and P. Legendre, and by the Canadian Scientific Submersible Facility. We are particularly grateful to L. J. Hamilton (backscatter analysis), S. Dray (R language and statistical analysis), M. Fellows (Imagenex surveys), Imagenex Corporation (post-processing support), and J. Illman (navigation software training) for advice and help. We also benefited from interesting comments by two anonymous reviewers.

---

## CHAPITRE 4

# CLASSIFICATION ET DÉTECTION D'HABITATS BENTHIQUES À L'AIDE DE SIGNATURES SONORES

---

Sébastien DURAND<sup>1</sup> and Pierre LEGENDRE<sup>1</sup>

Comptes rendus des 12èmes Rencontres de la Société Francophone de Classification,  
Montréal, 30 mai - 1er juin 2005.

P. 130-134 in : Makarenkov, V., G. Cucumel et F.-J. Lapointe [éds].

<sup>1</sup> *Département des Sciences Biologiques - Université de Montréal, C.P. 6128 succ.*

*Centre-Ville-Montréal (Québec), H3C-3J7, Canada E-mail :*



## 4.1 Résumé

Étudier la nature des grands fonds marins sur de vastes étendues a toujours été une tâche fastidieuse même en utilisant un sous-marin muni de lampes puissantes et de caméras. L'acquisition de données visuelles détaillées est limitée à de petites superficies à cause du manque de visibilité. Grâce à la télédétection sonore, l'étude spatiotemporelle des milieux benthiques profonds est dorénavant à la portée des écologistes. Nous avons élaboré une méthode d'analyse permettant de classifier les ondes sonores et de relier les groupes ainsi formés à des types d'habitats. Nous décrirons en détail certains aspects de la méthode : d'abord comment identifier les variables les plus représentatives des signaux sonores, puis déterminer l'influence de différentes variables du substrat marin sur les signaux sonores enregistrés.

MOTS-CLÉS : *Analyse discriminante canonique; écho sonore; habitat; radiale Juan de Fuca; sonar; source hydrothermale; submersible téléguidé; télédétection.*

## 4.2 Introduction

Depuis longtemps, l'homme tente de comprendre la nature des ondes sonores. Comme le font plusieurs autres espèces animales, nous utilisons le son afin de communiquer entre nous. Chez certains mammifères, le son sert aussi à des fins de navigation et même à l'identification d'objets éloignés. Le fond des abysses océaniques, composé de roc et de sédiments, baigne dans un milieu qui absorbe fortement la lumière. On y trouve des communautés animales et bactériennes qui vivent dans l'obscurité. L'aptitude des chauves-souris et des dauphins à détecter des objets à distance grâce à des échos sonores nous montre que les diverses textures et densités propres aux aires échantillonnées affectent d'une façon significative la nature des échos sonores qui peuvent être enregistrés par nos appareils. Par l'analyse de variables décrivant la forme et l'intensité de ces signaux, nous chercherons à identifier à distance les habitats d'un champ hydrothermal et, éventuellement, à cartographier les habitats du milieu benthique sur de grandes surfaces.

Il n'existe pas de méthode standard pour l'extraction, la transformation et le traitement des variables tirées de signaux sonores. Cela n'est pas surprenant. D'une part, les méthodes d'interprétation d'ondes sonores ne sont encore qu'en phase exploratoire ou sont gardées secrètes par des compagnies privées. D'autre part, de nombreux types d'habitats marins n'ont pas encore été étudiés et de nombreuses fréquences et tailles d'empreintes sonores n'ont pas encore été utilisées. Il serait donc difficile d'établir des standards pour une méthodologie aussi jeune. Le présent article constitue une introduction à l'approche analytique de ce nouveau champ d'études. Il décrit la méthode d'analyse des échos sonores que nous avons développée et utilisée dans nos travaux.

Au cours de nos travaux récents (Durand et al., 2002, 2005) basés sur ceux de Clarke et Hamilton (Clarke and Hamilton, 1999), nous avons [1] testé diverses méthodes de transformation des données ; [2] démontré visuellement l'importance de normaliser les données sonores en fonction de leur altitude d'acquisition (distance au fond marin) ; [3]

démontré que la variation de la taille des empreintes sonores influe sur les capacités discriminantes des ondes sonores ; et [4] créé un jeu de variables permettant la discrimination efficace des signatures sonores acquises dans les habitats dominants qui se retrouvent à l'intérieur de notre site d'étude.

## 4.3 Méthodes

### 4.3.1 Acquisition des données

Les données analysées dans cet article ont été récoltées par le ROPOS, un sous-marin canadien télécommandé depuis un navire de la garde côtière canadienne, le John P. Tully. Au cours de la mission *High Rise* en mai 2001, nous avons étudié le champ de sources hydrothermales Clambed situé par 2200 m de fond sur la radiale Juan de Fuca, à quelque 200 miles marins à l'ouest de l'île de Vancouver dans le Pacifique. Les caméras du ROPOS ont filmé les habitats de ce champ hydrothermal en même temps que nous récoltions des signaux sonores à l'aide d'un sonar Imagenex 881B. Le plan d'échantillonnage comportait 18 transects verticaux (1 à 10 m d'altitude) visant les cinq communautés principales ainsi que 12 transects horizontaux (1 à 5 m d'altitude) couvrant les principaux gradients naturels de ces milieux profonds. La fréquence d'échantillonnage moyenne était de trois signaux sonores par seconde.

### 4.3.2 Traitement des données et extraction des variables

Puisque les ondes sonores sont affectées par des phénomènes aléatoires comme le bruit ambiant, l'instabilité des capteurs du ROPOS et du sonar, ainsi que par la variabilité naturelle des signaux, le calcul de la moyenne de plusieurs échos successifs augmente la stabilité du signal sonore (Hamilton et al., 1999) et permet aussi l'obtention de données représentant un intervalle temporel choisi par le chercheur. Après avoir calculé la moyenne des signaux, on extrait une première variable, l'altitude, car c'est à partir de cette variable que nous effectuerons tous les filtrages initiaux, corrections et normalisation pour la profondeur. Puisqu'il n'existe aucune méthode standard d'extraction de la variable altitude, nous avons créé notre propre algorithme. Ainsi, à l'intérieur d'un écho sonore, nous avons utilisé comme marqueur temporel servant à l'estimation de l'altitude d'acquisition le point correspondant à 80 % de la valeur d'intensité maximum de l'onde sonore temporairement lissée.

Afin de permettre la comparaison des informations extraites de différents signaux sonores, une normalisation en fonction de l'altitude est nécessaire. En fait, lorsqu'une onde sonore se déplace dans l'eau, son intensité diminue, car non seulement l'onde s'étire de façon sphérique, mais l'eau servant de support absorbe aussi une partie de l'énergie de l'onde. Dans la plupart des cas, le taux d'absorption exacte au site d'échantillonnage n'est pas connu. Il est donc nécessaire d'en ajuster la valeur. Dans ce dessein, nous avons dessiné tous les signaux sonores côte à côte et attribué des niveaux de gris aux valeurs d'intensité qui les composent, puis nous avons comparé visuellement les signaux sonores normalisés acquis à des altitudes différentes. Nous avons modifié les valeurs de taux d'absorption jusqu'à ce que les niveaux de gris attribués aux signaux obtenus à différentes altitudes soient uniformes. Nous avons alors considéré que nous avons ajusté les taux d'absorption sonore.

Pour qu'un échantillon sonore soit valide et utilisé dans nos analyses, le deuxième écho doit être enregistré en entier. Cette filtration initiale garantit que toutes les variables, servant à décrire le premier et le deuxième écho, ne seront pas biaisées par des échos incomplets. Puisqu'aucun standard industriel n'indique quelles variables permettent une description optimale d'un signal sonore, nous avons dans un premier temps basé notre choix sur des variables facilement interprétables et, dans un deuxième temps, tenté de conserver les variables les plus différentes les unes des autres. Deux jeux de variables ont été utilisés et testés. Le premier décrit principalement la forme du premier et du deuxième écho par l'entremise de variables extraites (VE) telles que les coefficients d'aplatissement et d'asymétrie, la distance entre les centroïdes de diverses sections de l'écho, l'aire sous la courbe, etc. Le deuxième jeu de données (variables d'intensité, VI) utilise la série temporelle d'intensités centrées réduites du premier écho créée lors de la numérisation du signal sonore.



Pour chacune des 28 variables extraites (VE), nous avons sélectionné, parmi 4 transformations possibles, celle qui produisait les distributions les moins asymétriques. La transformation retenue différait d'une variable à l'autre. Pour l'ensemble des 92 variables d'intensité (VI), une seule transformation a été retenue parmi 11 transformations testées (Durand et al., 2005) pour transformer les variables  $k_j$  en  $k_j'$  :

$$p_j = k_j^{0.25} / \max(k^{0.25})$$

$$k_j' = \arcsinus(p_j^{0.5})$$

$(k_j^{0.25})$  représente la racine quatrième des valeurs d'une variable d'intensité ( $k_j$ ) décrivant un écho, suivie d'une transformation en proportion ( $p_j$ ) par rapport à l'intensité maximale du signal  $\max(k^{0.25})$ , puis de la transformation  $\arcsinus(p_j^{0.5})$  (Sokal and Rohlf, 1995). Puisque nous avons divisé l'intensité des signaux sonores par leur valeur maximale, cette transformation ne conserve que la forme des échos comme source de variation.

## 4.4 Analyses et résultats

### 4.4.1 Transects verticaux

Afin d'évaluer le pouvoir de discrimination des variables formant les deux jeux de données (VI et VE), nous avons utilisé les enregistrements vidéo pour attribuer à chacun des signaux sonores un nom d'habitat parmi les cinq habitats dominants observés au site hydrothermal Clambed. Nous avons supprimé toutes les variables sonores ayant une variance intragroupe nulle ou une trop faible variance.

L'analyse des jeux de variables VI, VE, ainsi que l'union de ces deux groupes, fut réalisée de la façon suivante. 70 % des échantillons sonores (5352 signaux), tirés au hasard dans chaque habitat, furent utilisés pour construire un modèle prédictif à l'aide d'analyses discriminantes linéaires. Le pouvoir discriminant de chaque modèle fut évalué en termes de pourcentage de classification correcte des données restantes, soit 30 % ou 2293 signaux. En comparant les pourcentages de classification correcte obtenus pour les trois jeux de données, nous avons noté que l'utilisation conjointe des deux jeux de données dans la même analyse (VI et VE) pouvait améliorer les prédictions de 10 %. Puis, utilisant seulement ce jeu de variables, les signaux sonores furent séparés en fonction de leur altitude d'acquisition afin d'évaluer les effets qu'ont les variations de la taille de l'empreinte sonore sur la capacité de discrimination. À l'aide des 5706, 1499 et 441 échantillons retrouvés respectivement entre 1 et 4, 4 et 7, et 7 et 10 mètres. Les pourcentages de classification correcte obtenus étaient faibles pour les basses altitudes (70.4 et 68.4 %) et plus élevés pour les signaux obtenus au-delà de 7 mètres (84 %. Fig. 4.1). Cette grande différence supporte l'hypothèse que les variations de la taille de l'empreinte peuvent grandement influencer les résultats d'un survol sonore.

Afin de faciliter l'interprétation des résultats, après sélection ascendante pas à pas parmi les 120 variables dont nous disposions au départ, nous avons retenu 16 variables

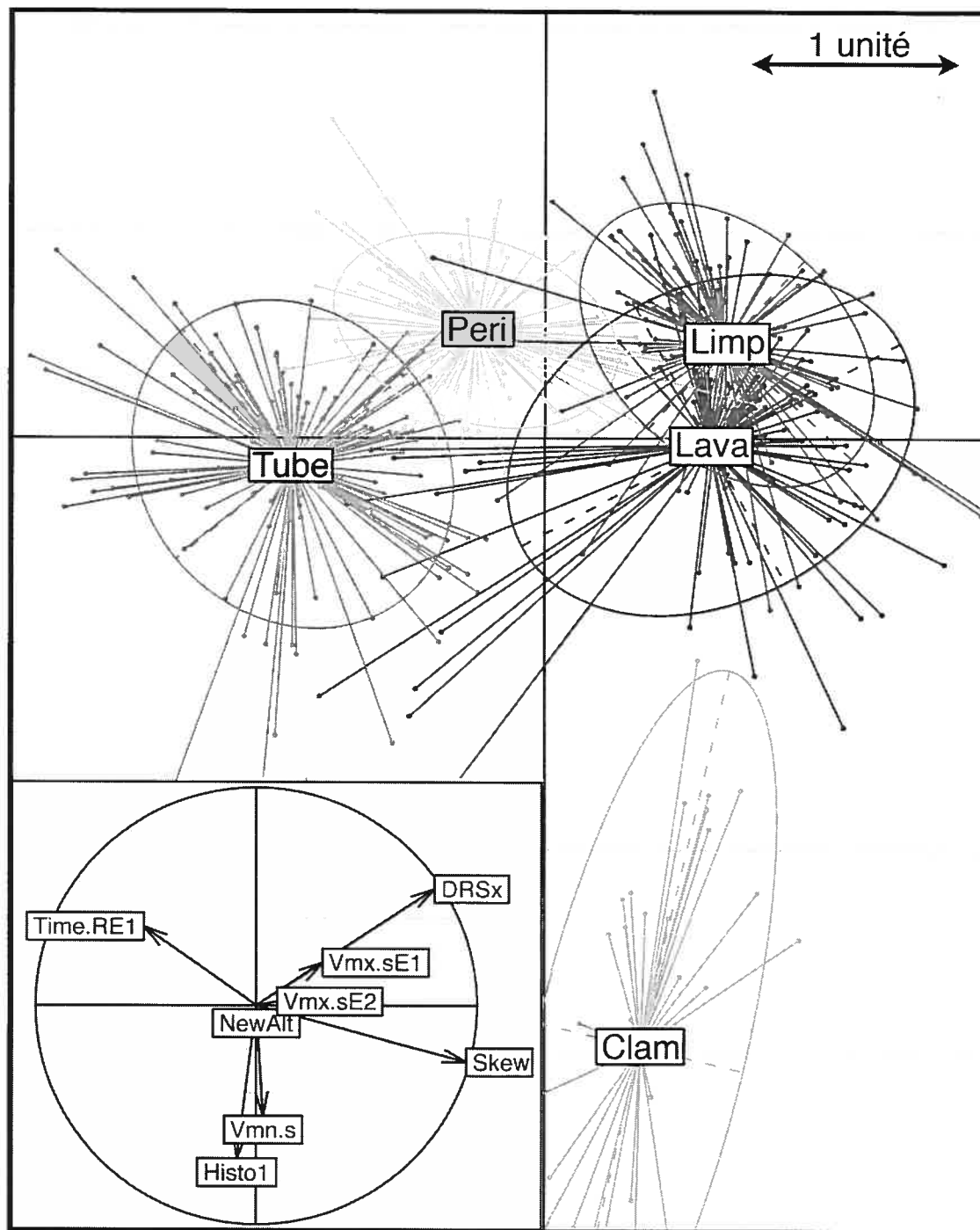


FIG. 4.1 - 30 % des échos échantillonnés à une altitude de 7 à 10 mètres projetés sur les deux premiers axes d'un modèle discriminant prédictif. Les variables VE dont les contributions aux axes (corrélations) sont illustrées en médaillon induisent la séparation des 5 habitats. L'espace discriminant est divisé en carrés de taille unité.

contribuant significativement à la discrimination. Malgré cette stricte sélection, les nouveaux pourcentages de classification correcte obtenus n'affichaient en moyenne qu'une baisse de 3 %. L'analyse plus approfondie des variables sélectionnées nous a permis non seulement d'identifier les caractéristiques propres à chaque signature sonore, mais aussi d'interpréter les signatures sonores en termes de texture et de densité.

#### 4.4.2 Transects horizontaux

Après avoir démontré les capacités discriminantes des ondes sonores, il nous restait à montrer la capacité réelle des données sonores pour l'identification correcte des différents types d'habitats lors de survols horizontaux du champ de sources hydrothermales. Puisque nous avons le projet d'identifier visuellement les signatures sonores enregistrées en vue de l'analyse canonique de redondance (ACR), les transects furent survolés à une altitude variant d'un à cinq mètres ; une altitude plus élevée n'aurait pas permis d'identifier les habitats avec suffisamment de précision sur les enregistrements vidéo. À cette distance, le cône sonore sur le fond marin formait une empreinte variant de trois à quinze centimètres de diamètre. Au cours des 12 transects parcourant le site d'étude, nous avons enregistré en simultanéité 13676 signaux sonores et 90 minutes de survol vidéo. Dans un premier temps, nous avons calculé la moyenne des échos regroupés par intervalles d'une seconde et avons réalisé une description visuelle de la nature de l'aire de l'empreinte sonore, seconde par seconde. Une fois les données transformées et normalisées pour la profondeur et les variables sonores extraites, nous avons tenté de découvrir à quelles caractéristiques visuelles les différentes variables sonores semblaient être les plus sensibles. Par sélection ascendante pas à pas, un sous-groupe de variables visuelles fut extrait, puis utilisé dans une analyse de redondance qui a permis d'identifier à quels éléments du paysage sous-marin correspondaient les différentes variables sonores. Parce que les données sonores contiennent beaucoup de bruit, le coefficient de redondance bimultivariable de l'ACR est faible ( $R^2 = 10.4 \%$ ). Il reste très hautement significatif ( $P = 0.001$  après 1000 permutations). Les relations entre les variables

visuelles et sonores illustrées à la Figure 4.2 sont en accord avec les relations habitats-variables sonores trouvées lors de l'analyse des transects d'altitude (Durand et al., 2005).

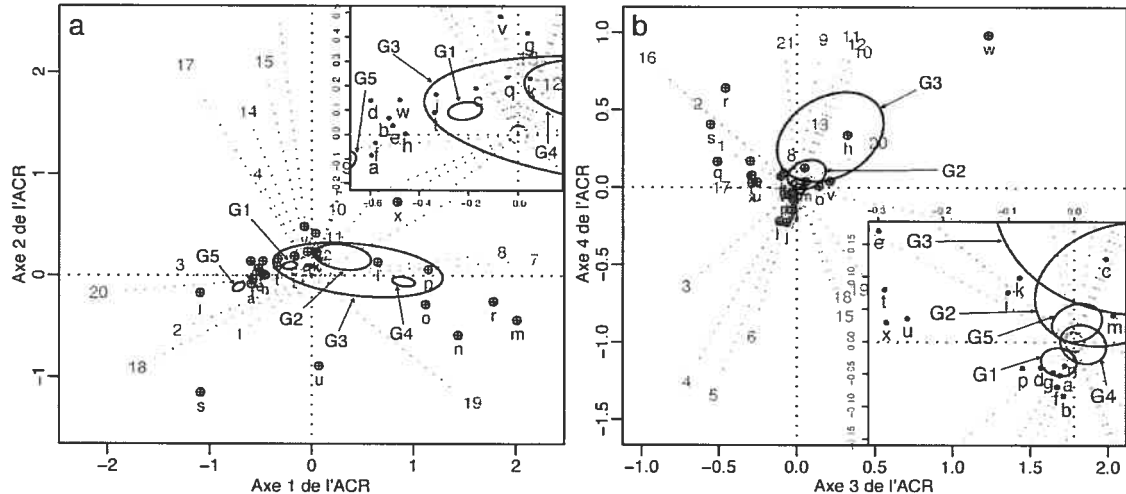


FIG. 4.2 – Projection des 21 variables sonores (variables VI : nos 1-13; variables VE : nos 14-21) et des 24 variables visuelles binaires (a à x) dans l'espace (a) des axes canoniques 1-2 et (b) des axes 3-4. Les cinq groupes (G1-G5) résultant de la partition par la méthode des  $K$  centroïdes sont représentés dans les graphiques par des ellipses de recouvrement à 95 %. à l'intérieur de chaque graphique, la zone centrale surchargée a été agrandie.

Nous avons ensuite partitionné les échos en 5 groupes de signaux sonores par la méthode des  $K$  centroïdes (*K-means*) (MacQueen, 1967) et retenu la solution présentant la plus faible variation intragroupe après 1000 démarrages aléatoires. Chaque groupe est représenté dans la Figure 4.2 par une ellipse de recouvrement à 95 %. Nous pouvons déterminer le type d'habitat représenté par chaque groupe (G1 à G5) en associant la position des ellipses aux variables visuelles (a à x) dans la figure. Cette étape était nécessaire pour nous permettre de cartographier les habitats détectés. Nos données de navigation contiennent cependant beaucoup d'erreur. Nous cherchons en ce moment à corriger ces données afin de pouvoir représenter sur une carte les différents groupes sonores que nous avons obtenus de la partition.

## 4.5 Discussion et conclusion

En nous basant sur les résultats obtenus au cours de cette étude exploratoire, nous croyons que l'utilisation du sonar à des fins de cartographie et de télédétection représente une avancée majeure pour l'étude des habitats benthiques. à ce jour, le plus grave problème relié à cette technologie est que les ondes enregistrées dans un même habitat sont très variables. De nombreuses combinaisons de variables à extraire des ondes sonores sont possibles ; elles n'ont pas encore été toutes créées ou testées. Afin de standardiser la méthode, nous devons chercher à identifier les combinaisons décrivant le plus adéquatement les divers échos sonores en vue de la cartographie des fonds marins. Parce que les basses fréquences pénètrent davantage le substrat, l'utilisation de plusieurs fréquences pourrait faire varier les signatures sonores provenant d'un même habitat et enrichir nos jeux de variables. Malgré l'augmentation du temps de calcul qui en résultera, l'utilisation conjointe de différentes fréquences sonores devrait permettre l'identification plus efficace des habitats benthiques, comme c'est le cas pour les différentes longueurs d'onde utilisées en télédétection terrestre.

Une des grandes sources de variance entre les échos est la variation de la taille de l'empreinte sonore sur le fond marin. Ces changements de taille de l'unité d'échantillonnage peuvent réduire grandement la capacité discriminante d'un jeu de données. Il est impératif que tout survol sonore respecte cette source de variation et tente de la minimiser. Ce ne sera qu'une fois cette méthode de télédétection bien établie et standardisée, que nous pourrons construire une banque de signatures sonores pouvant servir à l'identification des types d'habitats. Cela minimiserait les coûts associés à la validation visuelle des données sonores enregistrées au cours d'une mission.

---

## CHAPITRE 5

### CONCLUSION

---

*«Vagabonder à la surface des océans est souvent source de sérénité et, parfois, permet de tutoyer ses rêves. S'y immerger, c'est s'ouvrir à son observation et à sa compréhension.»*

Nicolas Hulot

Comme nous l'ont montré les dauphins et les chauves-souris, l'utilisation coordonnée du son et de l'ouïe peut permettre d'identifier la nature des objets qui nous entourent. La physiologie humaine ne permet pas un tel exploit. En développant le sonar, l'Homme a appris à émettre, recevoir et enregistrer des ondes sonores de hautes fréquences inaudibles à son oreille. Il s'agit d'une extension tout à fait synthétique de l'oreille humaine. L'analyse de ces signaux à l'aide d'ordinateurs nous permet de catégoriser les variations détectées afin de mieux différencier et reconnaître les objets rencontrés.

L'aspect exploratoire des trois articles présentés dans ce mémoire est dû au fait que l'utilisation du son à des fins de télédétection est une méthode encore nouvelle et d'une grande complexité. Plusieurs facteurs influencent l'information retrouvée dans les échos sonores. Premièrement, au moment de la réflexion sonore, les variables de forme et d'intensité d'un écho sont influencées par plusieurs facteurs tels que la texture, la densité et le relief de la surface échantillonnée. Deuxièmement, malgré l'excellente propagation du son dans l'eau, l'énergie d'un signal sonore est absorbée en fonction de la distance parcourue et dissipée dans l'eau. Finalement, au moment même de l'émission d'une onde sonore, de nombreux autres facteurs s'ajoutent à l'équation, tels que la fréquence d'émission, la longueur de pulsation ainsi que l'angle d'ouverture du faisceau sonore.

Un des facteurs qui influence le plus l'intensité et la forme des échos enregistrés est l'altitude d'acquisition. Or, afin d'optimiser la détection des textures et des densités propres aux sites échantillonnés, les échos doivent être normalisés pour la profondeur ou l'altitude d'acquisition. Cette normalisation est constituée de corrections en puissance et d'une correction temporelle. La correction de puissance corrige les échos pour l'absorption et la dissipation sonore, alors que la correction temporelle étire ou compresse les échos en fonction de la taille de l'empreinte sonore. Cette étape associée au traitement d'échos sonores est capitale. À l'intérieur de ce mémoire, à la figure 3.4, nous avons présenté une méthode graphique qui, par la visualisation des corrections de puissances, offre la possibilité d'ajuster et d'obtenir une normalisation optimale des signaux sonores.

Parmi tous les facteurs qui influencent la nature des échos sonores, l'effet de l'absorption et de la dissipation sonore peut être rectifié, mais celui des variations de la taille des empreintes sonores reste inaltérable. Puisqu'un faisceau sonore émit par un sonar est de forme conique et que son ouverture est constante, la taille de l'empreinte est proportionnelle à la distance qui sépare l'émetteur de la surface ciblée. Lorsque de grandes fluctuations d'altitude sont observées, les échos sonores échantillonnés représentent des échelles spatiales (taille de l'unité d'échantillonnage) différentes, ce qui rend la comparaison entre les échos impossible. Pour résoudre ce problème, il y a deux solutions. La première consisterait à utiliser un sonar ayant la capacité de conserver une taille constante d'empreinte sonore, mais ce dernier n'existe pas. La deuxième est simplement d'effectuer les survols à altitude constante, ce qui peut s'avérer problématique dépendamment des outils disponibles. Cependant, dans ce mémoire, nous avons plutôt tenté de tirer avantage de la conicité du faisceau sonore. À l'aide de transects verticaux que nous avons segmentés en trois classes d'altitude, soit les classes de 1-4, 4-7 et 7-10 m d'altitude, nous avons évalué l'influence de la taille de l'empreinte sonore. Pour chacune de ces classes, le diamètre de l'empreinte varie respectivement entre 3-12, 12-20 et 20-30 cm. À la suite de ce fractionnement, un fait très intéressant a été observé : avec



l'utilisation d'empreintes sonores plus larges, nous avons obtenu une augmentation de la capacité discriminante des signaux sonores. Ce phénomène souligne que l'utilisation d'empreintes sonores capturant un échantillon représentatif des variations de texture et de densité inhérentes à chaque habitat est une bonne façon d'optimiser les capacités discriminantes des survols sonores subséquents. Ainsi, grâce à de simples études préalables qui ne requièrent que quelques transects verticaux décrivant chacun des types d'habitats ciblés, il est possible d'optimiser rapidement des survols horizontaux qui seront effectués ultérieurement.

Au-delà des problèmes techniques associés à cette méthode, son utilisation comme moyen de télédétection représente d'ores et déjà une percée majeure pour l'étude des habitats benthiques. Lorsqu'on la compare aux méthodes vidéo, la méthode de télédétection sonore semble plus apte à décrire spatiotemporellement un site. Dans un premier temps, cette méthode permet non seulement l'acquisition de données bathymétriques, mais aussi et surtout l'acquisition d'information relative aux textures et aux densités qui sont propres aux habitats et substrats survolés. Deuxièmement, à l'aide de cette méthode, l'analyse des données sonores se fait automatiquement, le traitement des données est beaucoup plus rapide et les erreurs d'interprétation causées par l'évaluation visuelle arbitraire des enregistrements vidéo sont éliminées. Troisièmement, en pleine obscurité 2000 mètres sous la surface de la mer, nous avons constaté que la méthode vidéo nécessite une bonne visibilité; elle requiert donc une grande proximité avec le substrat échantillonné. Contrairement à cette méthode, l'acquisition de données sonores peut être effectuée en pleine noirceur et donc à plus grande distance du fond. Cette aptitude minimise de beaucoup les coûts d'exploration, car loin des reliefs escarpés associés aux dorsales médio-océaniques, les risques d'accident et les manoeuvres de contournement sont diminués. Malgré son jeune âge, cette méthode tant par son efficacité que par son potentiel ne demande qu'à être approfondie.

Lorsque les informations sonores sont intégrées à l'intérieur d'un système d'information géographique (SIG), la description spatiale des propriétés du substrat peut être

conjuguée à une bathymétrie extraite de ces mêmes signaux sonores. Si les données sont géoréférencées, l'association des informations topographiques aux informations de densité et de texture améliore notre capacité de répondre aux questions d'ordre écologique. Ce mémoire a réalisé un pas dans cette direction en montrant qu'à partir d'une seule fréquence sonore, il est possible d'identifier des habitats abyssaux. Considérant le fait que la pénétration et la sensibilité des signaux sonores varient en fonction des fréquences utilisées, il semble probable qu'une description plus exhaustive de la nature et de la composition des substrats échantillonnés soit possible par l'utilisation simultanée de différentes fréquences sonores.

Lorsque la sensibilité des ondes sonores aura été pleinement démontrée, l'essor de la méthode de télédétection sonore sera directement relié à l'indépendance qu'elle aura acquise des contrôles vidéo qui servent à l'identification des signatures sonores. Nous espérons donc que, dans un futur rapproché, les chercheurs réaliseront des études comparatives portant sur différentes fréquences sonores ainsi que divers jeux de variables dérivées des signaux sonores. Ces études pourront servir de bases à l'élaboration d'une banque de signatures sonores ou d'une charte d'identification des échos sonores. De tels jeux de données permettront ultimement l'étude spatiotemporelle du plus méconnu et du plus vaste écosystème de notre planète, celui des grands fonds océaniques.

## BIBLIOGRAPHIE

- C.J. Allegre and S.H. Schneider. The evolution of the earth. *Scientific American*, 271 (4) :66, 1994.
- R.D. Ballard. Notes on a major oceanographic find. *Oceanus*, 20(3) :35–44, 1977.
- N.J. Bax, R.J. Kloser, A. Williams, K. Gowlett-Holmes, and T. Ryan. Seafloor habitat definition for spatial management in fisheries : a case study on the continental shelf of southeast australia. *Oceanologica Acta*, 22(6) :705–720, 1999.
- E. Blöchl, R. Rachel, S. Burggraf, D. Hafenbradl, H.W. Jannasch, and K.O. Stetter. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees c. *Extremophiles*, 1 :14–21, 1997.
- B. Brasier, O. Green, J. Lindsay, and A. Steele. Earth's oldest ( 3.5 ga) fossils and the 'early eden hypothesis' : questioning the evidence. *Origins of Life and Evolution of the Biosphere*, 34 :257–269, 2004.
- C.J. Brown, K.M. Cooper, W.J. Meadows, D.S. Limpenny, and H.L. Rees. Small-scale mapping of sea-bed assemblages in the eastern english channel using sidescan sonar and remote sampling techniques. *Estuarine, Coastal and Shelf Science*, 54 :263–278, 2002.
- J. Burczynski. Bottom classification, 2001.
- D.R. Burns, C.B. Queen, H. Sisk, W. Mullarkey, and R.C. Chivers. Rapid and convenient acoustic sea-bed discrimination for fisheries applications. *Proceedings of the Institute of Acoustics*, 11(3) :169–178, 1989.
- D.A. Carey, D.C. Rhoads, and B. Hecker. Use of laser line scan for assessment of response of benthic habitats and demersal fish to seafloor disturbance. *Journal of Experimental Marine Biology and Ecology*, 285-286 :435–452, 2003.

- R.C. Chivers, N. Emerson, and D.R. Burns. New acoustic processing for underway surveying. *The hydrographic journal*, 56 :9–17, 1990.
- P.A. Clarke and L.J. Hamilton. The abcs program for the analysis of echo sounder returns for acoustic bottom classification. Technical Report DSTO-GD-0215, Defence Science and Technology Organisation, 1999.
- J.B. Corliss and R.D. Ballard. Oases of life in the cold abyss. *National Geographic*, 1977(October) :441–453, 1977.
- J.R. Delaney, V. Robigou, R.E. McDuff, and M.K. Tivey. Geology of a vigorous hydrothermal system on the endeavour segment, juan de fuca. *Journal of Geophysical Research*, 97(B13) :19663–19682, 1992.
- S. Durand, M. Le Bel, S.K. Juniper, and P. Legendre. The use of video surveys, a geographic information system and sonar backscatter data to study faunal community dynamics at juan de fuca ridge hydrothermal vents. *Cahiers de Biologie Marine*, 43 : 235–240, 2002.
- S. Durand, P. Legendre, and S.K. Juniper. Sonar backscatter differentiation of dominant macrohabitat types in a hydrothermal vent field. *Ecological Applications*, accepted, 2005.
- K. E. Ellingsen, J.S. Gray, and E. Bjørnbom. Acoustic classification of seabed habitats using the qtc view system. *Journal of Marine Sciences*, 59 :825–835, 2002.
- R.L. Foster-Smith and I.S. Sotheran. Mapping marine benthic biotopes using acoustic ground discrimination systems. *International Journal Remote Sensing*, 24(13) :2761–2784, 2003.
- J.F. Grassle. The ecology of deep sea hydrothermal vent communities. *Advances in Marine Biology*, (23) :301–362, 1986.

- A. Grehan and S.K. Juniper. Clam distribution and subsurface hydrothermal processes at chowder hill (middle valley), juan de fuca ridge. *Marine Ecology Progress Series*, 130 :105–115, 1996.
- L.J. Hamilton. Acoustic seabed classification systems. Technical Report DSTO-TN-0401, Defense Science & Technology Organisation, 2001.
- L.J. Hamilton, P.J. Mulhearn, and R. Poeckert. Comparison of roxann and qtc-view acoustic bottom classification system performance for the cairns area, great barrier reef, australia. *Continental Shelf Research*, 19 :1577–1597, 1999.
- J.E. Hewitt, S.F. Thrush, P. Legendre, G.A. Funnell, J. Ellis, and M. Morrison. Mapping of marine soft-sediment communities : Integrated sampling for ecological interpretation. *Ecological Applications*, 14(4) :1203–1216, 2004.
- J.L. Irish and W.J. Lillycrop. Scanning laser mapping of the coastal zone : the shoals system. *Journal of Photogrammetry & Remote Sensing*, 54 :123–129, 1999.
- H.W. Jannasch and M.J. Mottl. Geomicrobiology of deep-sea hydrothermal vents. *Science*, 229 :717–725, 1985.
- S.K. Juniper. Deep-sea hydrothermal vent and seep habitats and related governance issues. Technical report, Workshop Secretariat for the High Seas Biodiversity Program and general conference enquiries, 2003.
- S.K. Juniper, J. Sarrazin, and A. Grehan. Remote sensing of organism density and biomass at hydrothermal vents. *Cahiers de Biologie Marine*, 39 :245–247, 1998.
- K. Kashefi and D.R. Lovley. Extending the upper temperature limit for life. *Science*, 301 :934, 2003.
- D.S. Kelley, J.A. Baross, and J.R. Delaney. Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annual Review of Earth and Planetary Sciences*, 30 :385–491, 2002.

- R.E. Kochevar and J.J. Childress. Carbonic anhydrase in deep-sea chemoautotrophic symbioses. *Marine Biology*, 125(2) :375–383, April 1996.
- S.T. Kotchenova, X. Song, N.V. Shabanov, C.S. Potter, Y. Knyazikhin, and R.B. Myrneni. Lidar remote sensing for modeling gross primary production of deciduous forests. *Remote Sensing of Environment*, 92 :158–172, 2004.
- K-H. Lee and E.N. Anagnostou. A combined passive/active microwave remote sensing approach for surface variable retrieval using tropical rainfall measuring mission observations. *Remote Sensing of Environment*, 92 :112–125, 2004.
- S.H. Lee and S.K. Chough. High-resolution (2-7 khz) acoustic and geometric characters of submarine creep deposits in the south korea plateau, east sea. *Sedimentology*, 48 : 629–644, 2001.
- L. Legendre and E.D. Gallagher. Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129 :271–280, 2001.
- P. Legendre, K. E. Ellingsen, E. Bjørnbom, and P. Casgrain. Acoustic seabed classification : improved statistical method. *Canadian Journal of Fisheries and Aquatic Sciences*, 59 :1085–1089, 2002.
- R.J. L evell e and S.K. Juniper. Biogeochemistry of deep-sea hydrothermal vents and cold seeps. In K.D. Black Shimmield and G.B., editors, *Biogeochemistry of Marine Systems*, pages 238–292. Blackwell, Sheffield, UK., 2003.
- J. MacQueen. Some methods for classification and analysis of multivariate observations. In Le Cam L.M. and Neyman J., editors, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Berkeley, 1967.
- R.G. Masson. A magnetic survey off the west coast of the united states between latitudes 32° and 36° n, longitudes 121° and 128° w. *Geophysical Journal*, 1 :320–329, 1958.

- I. Moya, L. Camenen, S. Evain, Y. Goulas, Z.G. Cerovic, G. Latouche, J. Flexas, and A. Ounis. A new instrument for passive remote sensing, 1. measurements of sunlight-induced chlorophyll fluorescence. *Remote Sensing of Environment*, 91 :186–197, 2004.
- D.C. Nelson, K.D. Hagen, and D.B. Edwards. The gill symbiont of the hydrothermal vent mussel *Bathymodiolus thermophilus* is a psychrophilic, chemoautotrophic, sulfur bacterium. *Marine Biology*, 121(3) :487–495, January 1995.
- E.G. Nisbet and N.H. Sleep. The habitat and nature of early life. *Nature*, 409 :1083–1091, February 2001.
- N.G. Pace and H. Gao. Swathe seabed classification. *Journal of Oceanic Engineering*, 13(2) :83–90, 1988.
- D.M. Parry, M.A. Kendall, D.A. Pilgrim, and M.B. Jones. Identification of patch structure within marine benthic landscapes using a remotely operated vehicle. *Journal of Experimental Marine Biology and Ecology*, 285-286 :497–511, 2003.
- B.T. Prager, D.A. Caughey, and R.H. Poeckert. Bottom classification : operational results from qtc view. In *OCEANS'95 - Challenges of Our Changing Global Environment*, volume 3, pages 1827–1835, San Diego, California, 1995.
- J. Sarrazin and S.K. Juniper. The use of video imagery to gather biological information at deep-sea hydrothermal vents. *Cahiers de Biologie Marine*, 39 :255–258, 1998.
- J. Sarrazin, V. Robigou, S.K. Juniper, and J.R. Delaney. Biological and geological dynamics over four years on a high-temperature sulfide structure at the Juan de Fuca ridge hydrothermal observatory. *Marine Ecology Progress Series*, 153 :5–24, 1997.
- J. Sarrazin, S.K. Juniper, G.J. Massoth, and P. Legendre. Physical and chemical factors influencing species distributions on hydrothermal sulfide edifices of the Juan de Fuca ridge, northeast Pacific. *Marine Ecology Progress Series*, 190 :89–112, 1999.

- S. Schmidtlein and J. Sassin. Mapping of continuous floristic gradients in grasslands using hyperspectral imagery. *Remote Sensing of Environment*, 92 :126–138, 2004.
- J.W. Schopf. *Cradle of life : The Discovery of Earth's Earliest Fossils*. Princeton University Press, 1999.
- R.R. Sokal and F.J. Rohlf. *Biometry*. W.H. Freeman, New York, third edition edition, 1995.
- C.J.F. ter Braak and P. Smilauer. Canoco reference manual and canodraw for windows user's guide : software for canonical community ordination (version 4.5)., 2002.
- M.K. Tivey, L.O. Olson, V.W. Miller, and R.D Light. Temperature measurements during initiation and growth of a black smoker chimney. *Nature*, 346(51-54), 1990.
- M. Tsurumi and V. Tunnicliffe. Tubeworm-associated communities at hydrothermal vents on the juan de fuca ridge, northeast pacific. *Deep-Sea Research*, 50 :611–629, 2003.
- V. Tunnicliffe. Observation on the effects of sampling on hydrothermal vent habitat and fauna of axial seamount, juan de fuca ridge. *Journal of Geophysical Research*, 95(B8) :12961–12966, 1990.
- V. Tunnicliffe. The nature and origin of the modern hydrothermal vent fauna. *Palaios*, 7 :338–350, 1992.
- V. Tunnicliffe and C.M.R. Fowler. Influence of sea-floor spreading on the global hydrothermal vent fauna. *Nature*, 379 :531–533, 1996.
- V. Tunnicliffe and R. Thomson. Un projet pilote de la zone de protection marine de l'océan pacifique canadien. Technical report. Pêches et Océans Canada, 1999.
- V. Tunnicliffe, R.W. Embley, J.F. Holden, D.A Butterfield, G.J. Massoth, and S.K. Juniper. Biological colonization of new hydrothermal vents following an eruption on juan de fuca ridge. *Deep-Sea Research*, 44(9-10) :1627–1644, 1997.



- F.J. Vine and D.H. Drummond, Matthews. Magnetic anomalies over oceanic ridges. *Nature*, CXCIX :1947–1949, 1963.
- W.S.D. Wilcock, S.D. Archer, and G.M. Purdy. Microearthquakes on the endeavour segment of the juan de fuca ridge. *Journal of Geophysical Research*, 107(B12) :2336, 2002.
- C.O. Wirsen, S.M. Sievert, C.M. Cavanaugh, S.J. Molyneaux, A. Ahmad, L.T. Taylor, E.F. DeLong, and C.D. Taylor. Characterization of an autotrophic sulfide oxidizing marine arcobacter sp that produces filamentous sulfur. *Applied and Environmental Microbiology*, 68 :316–325, 2002.
- R.N. Zajac, R.S. Lewis, L.J. Poppe, D.C. Twichell, J. Vozarik, and M.L. DiGiacomo-Cohen. Responses of infaunal populations to benthoscape structure and the potential importance of transition zones. *Limnology and Oceanography*, 48(2) :829–842, 2003.

---

## APPENDICE A

**THE DEFINITIONS OF THE VE VARIABLES ALONG  
WITH THEIR TRANSFORMATION DETAILS, THE  
TRANSFORMATION TRIALS FOR THE VI DATA SETS,  
AND THE VARIABLES RETAINED BY FORWARD  
SELECTION ARE SHOWN.**

---

TAB. A.I: Definitions, transformations, and skewness values of the VE variables.

Variable name	Definition	Transformation	Skewness
NewAlt	The computed altitude	Double Sqrt	1.29
Pmx.sE1	Point of maximum found in the smoothed E1	Double Sqrt	1.24
Pmn.s	Point of minimum found between smoothed E1 and E2	Sqrt	1.26
Pmx.sE2	Point of maximum found in the smoothed E2	Double Sqrt	1.36
Vmx.sE1	Maximum value in the smoothed E1	None	0.09
Vmn.s	Minimum value found between smoothed E1 and E2	Double Sqrt	3.48
Vmx.sE2	Maximum value in the smoothed E2	Double Sqrt	4.18
DRSx	Time distance between rise and set area centroids of E1	None	0.72
DRSy	Intensity distance between rise and set area centroids of E1	None	0.14
Area.R	Area below the curve of E1 rise section	Double Sqrt	1.60
Area.S	Area below the curve of E1 set section	Double Sqrt	1.48
Area.E2	Area below the curve of E2 rise section	Double Sqrt	11.78
Area.RE1	Proportion between E1 rise and E1 total area	None	0.10
Area.RE2	Proportion between E1 rise and E2 rise area	Double Sqrt	49.80
Time.R	Time laps of E1 rise section	Sqrt	1.15
Time.S	Time laps of E1 set section	Double Sqrt	2.06
Time.E2	Time laps of E2 rise section	Double Sqrt	4.79
Time.RE1	Proportion between E1 rise and E1 total time laps	Log	0.26
Time.RE2	Proportion between E1 rise and E2 rise time laps	Double Sqrt	6.72
Var	E1 variance (second statistical moment)	Double Sqrt	2.25

Continued on next page

Tab. A.I – continued from previous page

Variable name	Definition	Transformation	Skewness
Skew	E1 skewness (derived from third statistical moment)	Log	0.20
Histo1	1st histogram class of E1 intensities, based on matrix max	Double Sqrt	1.60
Histo2	2nd histogram class of E1 intensities, based on matrix max	Double Sqrt	1.07
Histo3	3rd histogram class of E1 intensities, based on matrix max	Log	0.58
Histo4	4th histogram class of E1 intensities, based on matrix max	Double Sqrt	2.28
Histo5	5th histogram class of E1 intensities, based on matrix max	Double Sqrt	13.94
Histo6	6th histogram class of E1 intensities, based on matrix max	Double Sqrt	32.86
Histo7	7th histogram class of E1 intensities, based on matrix max	Double Sqrt	69.48

*Note :* The VE variables used in our analyses were selected out of a larger set of 66 variables in such way as to minimize the collinearity among the variables. Then, out of four transformations (none, log, sqrt, and double sqrt), the transformation producing the lowest skewness values was applied to each of these variables. For all rising, setting and complete sections found in our backscatters, the areas under the curves and time spent were calculated. The following area and time spent proportions were also calculated : E1 rise / E1 complete and E1 rise / E2 rise. The points and values of maximum intensity found in E1 and E2 on both the original and smoothed curves and the point and value of minimum intensity found between these maxima were also described. Finally, the statistical moments, a 7-class histogram of E1 intensity values, and the distances, in the x and y directions, between the centroids of the E1 rising and setting sections were calculated.

TAB. A.II: Transformation trial for the VI data table and subtables.

Table types	First transformation	Second transformation		Table types	First transformation	Second transformation		
		None	Arcsine			None	Arcsine	None
STB	None	2.05	1.30	1.05	None	10.62	7.70	7.49
	Sqrt	1.07	0.69	0.57	Sqrt	7.51	6.06	5.82
	Double Sqrt	0.69	<b>0.44</b>	0.62	Double Sqrt	5.95	<b>5.51</b>	5.58
	Log	1.93	—	—	Log	10.51	—	—
	Mod. Arcsine	1.13	—	—	Mod. Arcsine	7.57	—	—
BEST	None	6.04	3.96	3.81	None	4.66	3.32	3.09
	Sqrt	3.82	2.75	2.61	Sqrt	3.14	2.58	2.48
	Double Sqrt	2.65	<b>2.40</b>	2.55	Double Sqrt	2.50	<b>2.37</b>	2.56
	Log	5.96	—	—	Log	4.61	—	—
	Mod. Arcsine	3.86	—	—	Mod. Arcsine	3.15	—	—
ALL	None	13.47	8.89	8.52	None	3.06	2.32	2.17
	Sqrt	8.79	6.66	6.31	Sqrt	2.19	1.87	1.70
	Double Sqrt	6.50	<b>5.99</b>	6.09	Double Sqrt	1.74	1.68	<b>1.66</b>
	Log	13.36	—	—	Log	3.03	—	—
	Mod. Arcsine	8.86	—	—	Mod. Arcsine	2.21	—	—

*Note* : The presented values are the means of all skewness values computed on all VI data tables after different sequences of transformations. The lowest (best) skewness value for each data subset is in bold. Illogical transformations were not calculated (—). In the first set of transformation, either no transformation (None) was applied, or the effect of high intensity outliers on the distribution was reduced through a square root (Sqrt), double square root (Double Sqrt), log (Log), or modified arcsine transformation (Mod.Arcsine). This modified form used the table maximum as the denominator in the calculation of proportions, prior to computing the square root and then the arcsine, instead of the object's maximum in the ordinary arcsine transformation (Sokal and Rohlf, 1995). With this modification, all these transformations conserved both the strength and shape information found in backscatters. In a second step, on all the newly transformed tables, we performed either no transformation (None), an ordinary arcsine transformation (Arcsine :  $\arcsin((y/y_{max})^{0.5})$ , (Sokal and Rohlf, 1995), or a Hellinger transformation (Hellinger : (Legendre and Gallagher, 2001)). The last two removed the overall intensity of the signal and preserved only the shape information of the backscatters. STB, BEST, 1-4, 4-7, and 7-10 are subtables of ALL which refers to the complete data set. The STB subtable contained only the echoes acquired while the visibility was optimal and the remotely operated vehicle was in a stable position, hovering at low altitude over the selected habitat. The BEST subtable contained only the best transect for each habitat type. The 1-4, 4-7, and 7-10 subtables contained echoes acquired at different altitude ranges, respectively 1 to 4, 4 to 7, and 7 to 10 m.

TAB. A.III: Variables selected by forward selection, making the FWD variable sets.

Method	Table type	Selected variables
VI	STB	Int46-50, Int26-30, Int81-85, Int41-45, Int71-75, Int1-5, Int11-15, Int31-35
	BEST	Int61-65, Int31-35, Int91-95, Int21-25
	ALL	Int51-55, Int71-75, Int26-30, Int41-45, Int11-15
	1 - 4 m	Int55-60, Int41-45, Int71-75, Int26-30, Int11-15
	4 - 7 m	Int51-55, Int26-30, Int171-175, Int6-10, Int101-105
	7 - 10 m	Int181-185, Int31-35, Int56-60, Int11-15, Int96-100, Int186-190, Int206-210
VE	STB	Skew, Vmn.s, NewAlt, Histo1, Pmn.s, Time.R, Time.RE1, Histo4, Time.S, DRSx
	BEST	DRSx, Vmx.sE1, Vmx.sE2, NewAlt, Var
	ALL	Skew, Vmn.s, Pmx.sE1, NewAlt, Vmx.sE2, DRSx
	1 - 4 m	DRSx, Vmn.s, Pmx.sE1, Vmx.sE2, Skew, NewAlt, Histo1
	4 - 7 m	Skew, Histo1, Vmx.sE2, Time.RE1, Pmx.sE2, DRSx, Pmn.s
	7 - 10 m	Histo1, Skew, DRSx, Vmx.sE1, Time.RE1, Time.R, Time.S
VT	STB	Int46-50, Int26-30, Vmn.s, Int71-75, NewAlt, Skew, Int1-5, Int41-45, Int81-85, DRSx, Int16-20, Pmn.s
	BEST	DRSx, Int31-35, Vmx.sE2, Vmx.sE1, NewAlt, Int91-95
	ALL	Int51-55, Vmn.s, DRSx, Vmx.sE1, Int71-75, Int41-45, Pmx.sE1, Histo2
	1 - 4 m	Int56-60, Vmn.s, Skew, Int41-45, Int21-25, Vmx.sE2, Int71-75, Pmx.sE1
	4 - 7 m	Skew, Int6-10, DRSx, Int26-30, Int171-175, NewAlt, Int106-110, Int51-55, Int21-25, Vmx.sE2, Vmn.s
	7 - 10 m	Int181-185, Skew, DRSx, Int31-35, Int71-75, Vmx.sE1, Time.R, Int291-295, Time.RE1

*Note* : The STB, BEST, 1-4, 4-7, and 7-10 subtables are described in the notes of Table A.II. Based on their appearance frequencies, the VE variables “DRSx”, “Histo1”, “Skew”, “time.RE1”, “Vmn.s”, “Vmx.sE1”, “Vmx.sE2”, “Newalt” were selected to be included in the SEL variable set. For the VI set, the distribution of the variables selected by forward selection showed that 20% of the variables were describing the rise area of the first echo, and 60% for the set area. Consequently, based on these ratios and on the variable appearance frequencies, “Int11-15”, “Int21-25”, “Int31-35”, “Int46-50”, “Int56-60”, “Int71-75”, “Int91-95”, and “Int181-185” were also selected to be part of set SEL.



---

APPENDICE B

**HABITAT CLUSTERS, OBTAINED FROM DIFFERENT  
ALTITUDE RANGES USING THE SEL VARIABLE  
SUBSET, AND PROJECTED ON THE FIRST AND  
SECOND DISCRIMINANT AXES.**

---

Plots of habitat cluster projections on the first and second discriminant axes. The SEL subset of variables was used to construct the six figures presented in this Appendix. The first three figures are based on the VI data tables and the following three on the VE data tables. VT graphics are not shown since the nature of their variable sets makes it harder to unravel the links and relationships between the variables and the centroids of the habitat clusters created by discrimination analysis (R-Pkg : :ade4 :Discrimin). In each figure, 4 panels (a to d) and one table (e) are shown. In panels (a) to (d), discriminant axis 1 is the abscissa and axis 2 is the ordinate. Panel (a) shows the canonical weights of the variables; panel (b) shows the variables at angles representing their correlations relative to the first two discriminant axes; in panel (c), the eigenvalues of 4 discriminant axes are given. Panel (d) shows the backscatter clusters based on their video-assigned habitat types; for each habitat cluster, the nametag is located at the cluster centroid. In panel (e), the classification table compares the habitats assigned by discriminant analysis to the video-assigned habitats. PCC is the percentage of correct classification. In the (d) panels shown in Figures B.1-B.6, the Limp and Clam habitats are always visually well separated from the other habitat clusters. The habitat assignments in the contingency tables of Figs. B.1, B.2, B.4, and B.5 also differentiate these two major groups. It is only in Figures B.3 and B.6, which describe echoes acquired between 7 and 10 m of altitude, that most echoes are correctly assigned.

All selected VI variables described in panels (a) and (b) of Figs. B.1 to B.3 have at one altitude range or another shown a high canonical weight or correlation with one of the axes. This behavior suggests that VI is a good variable set for habitat classification. When the VE variables are used (Figs. B.4-B.6), the “Skew” and “DRSx” variables, which always have strong correlations and canonical weights with the axes, played influential roles on our discrimination results. They allowed us to identify the sonar signatures of the five dominant habitats found in a hydrothermal vent field, 2200 m below the sea surface.

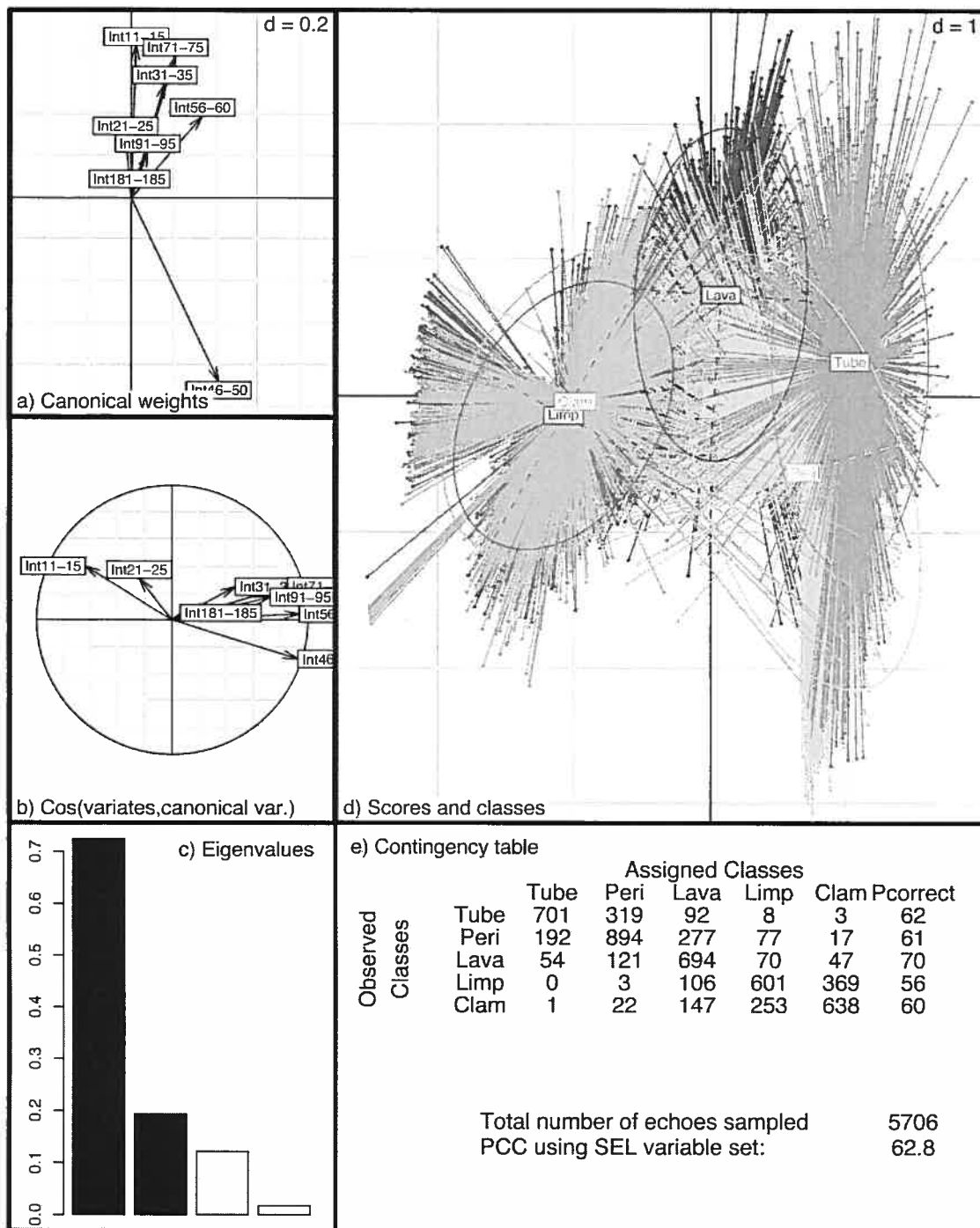


FIG. B.1 – Backscatter discrimination results using VI variables and subtable 1-4m

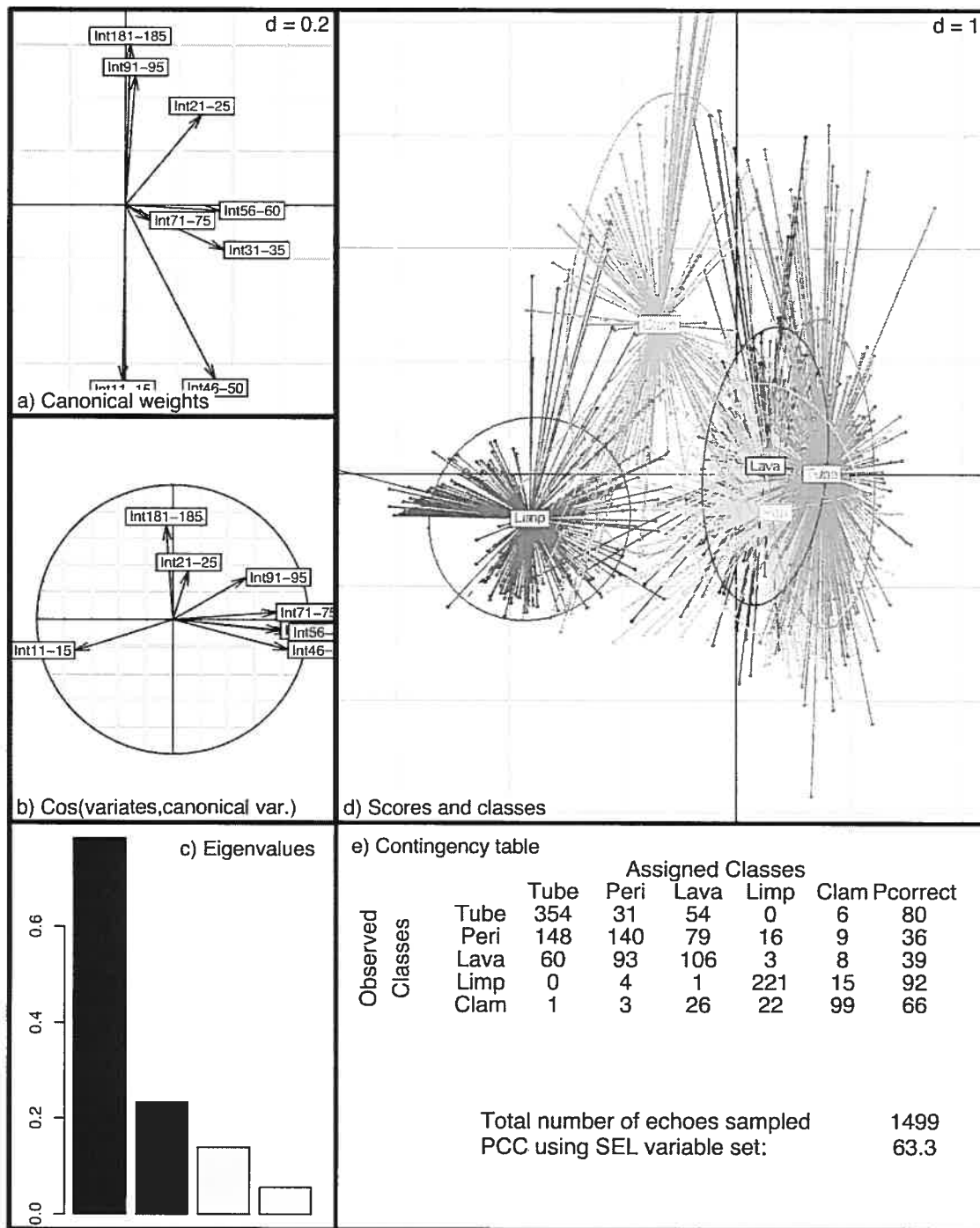


FIG. B.2 – Backscatter discrimination results using VI variables and subtable 4-7m

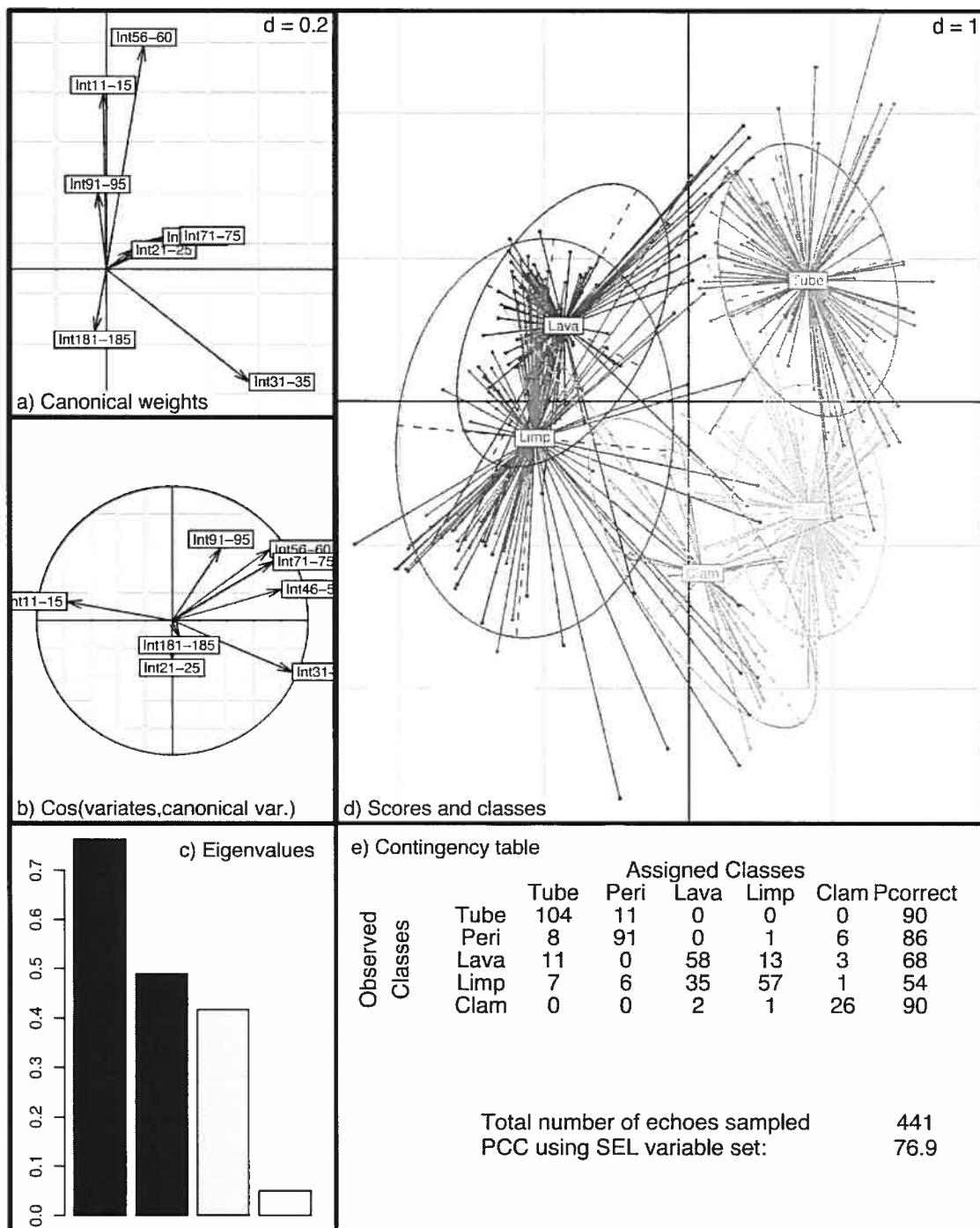


FIG. B.3 – Backscatter discrimination results using VI variables and subtable 7-10m

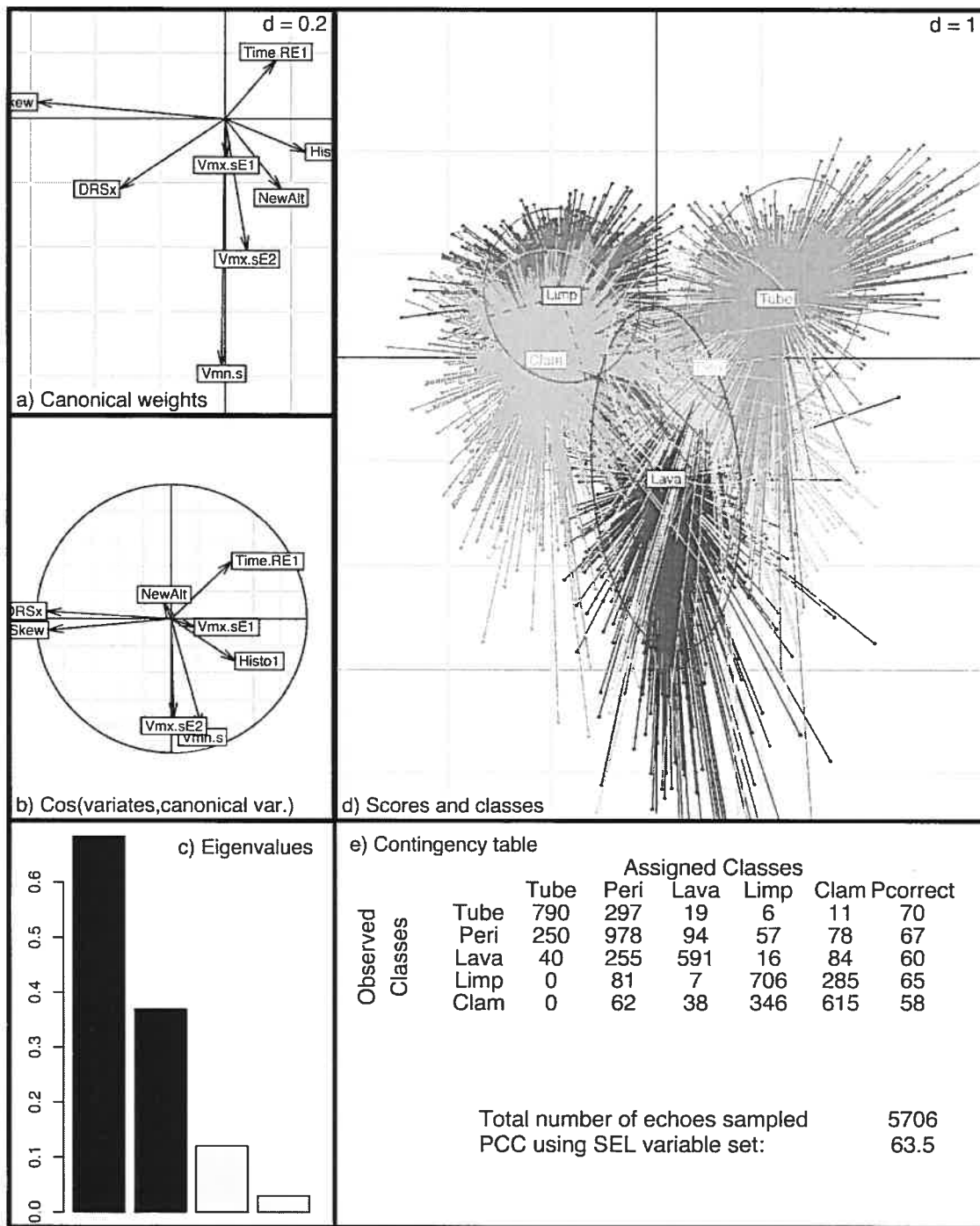


FIG. B.4 – Backscatter discrimination results using VE variables and subtable 1-4 m

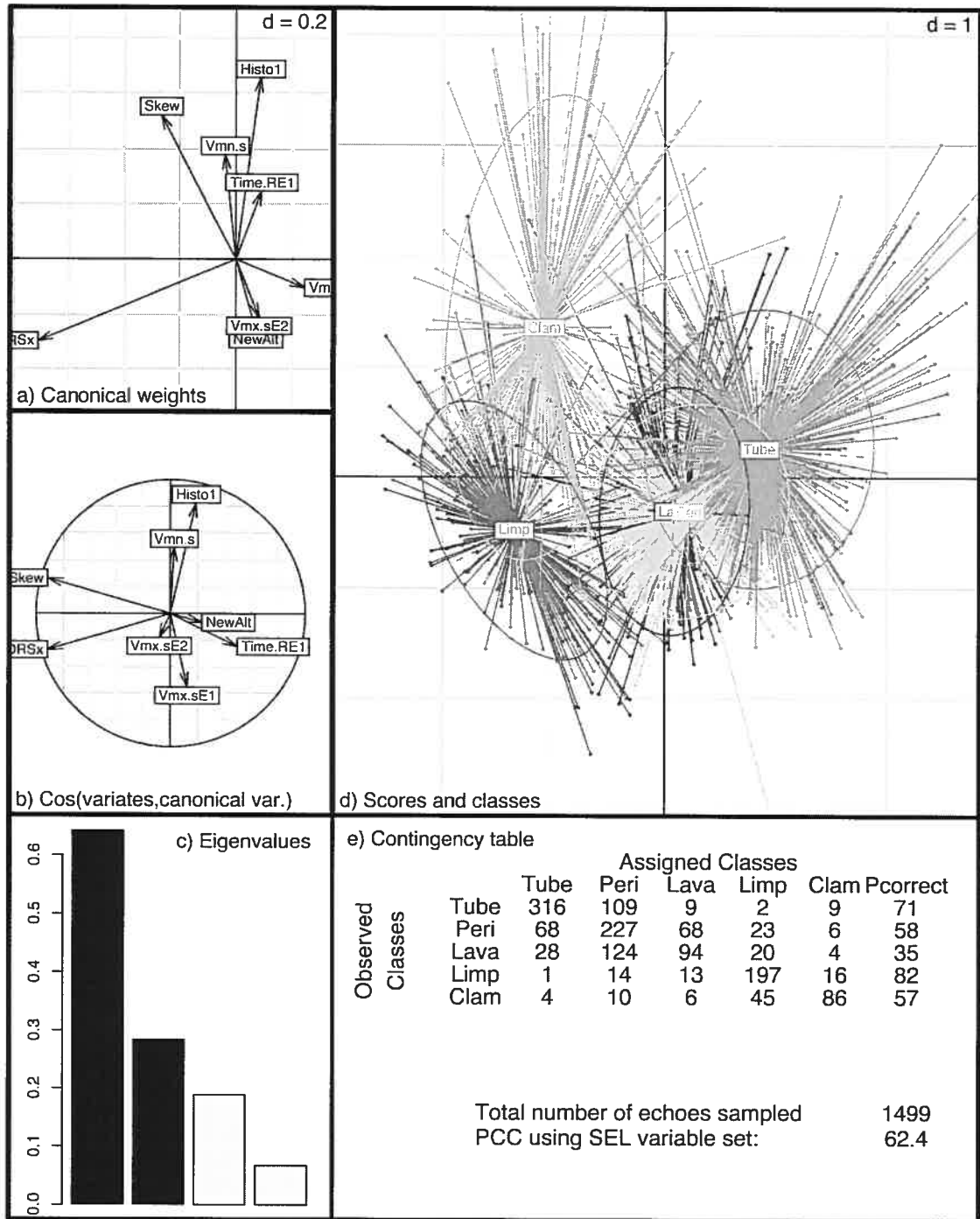


FIG. B.5 – Backscatter discrimination results using VE variables and subtable 4-7 m

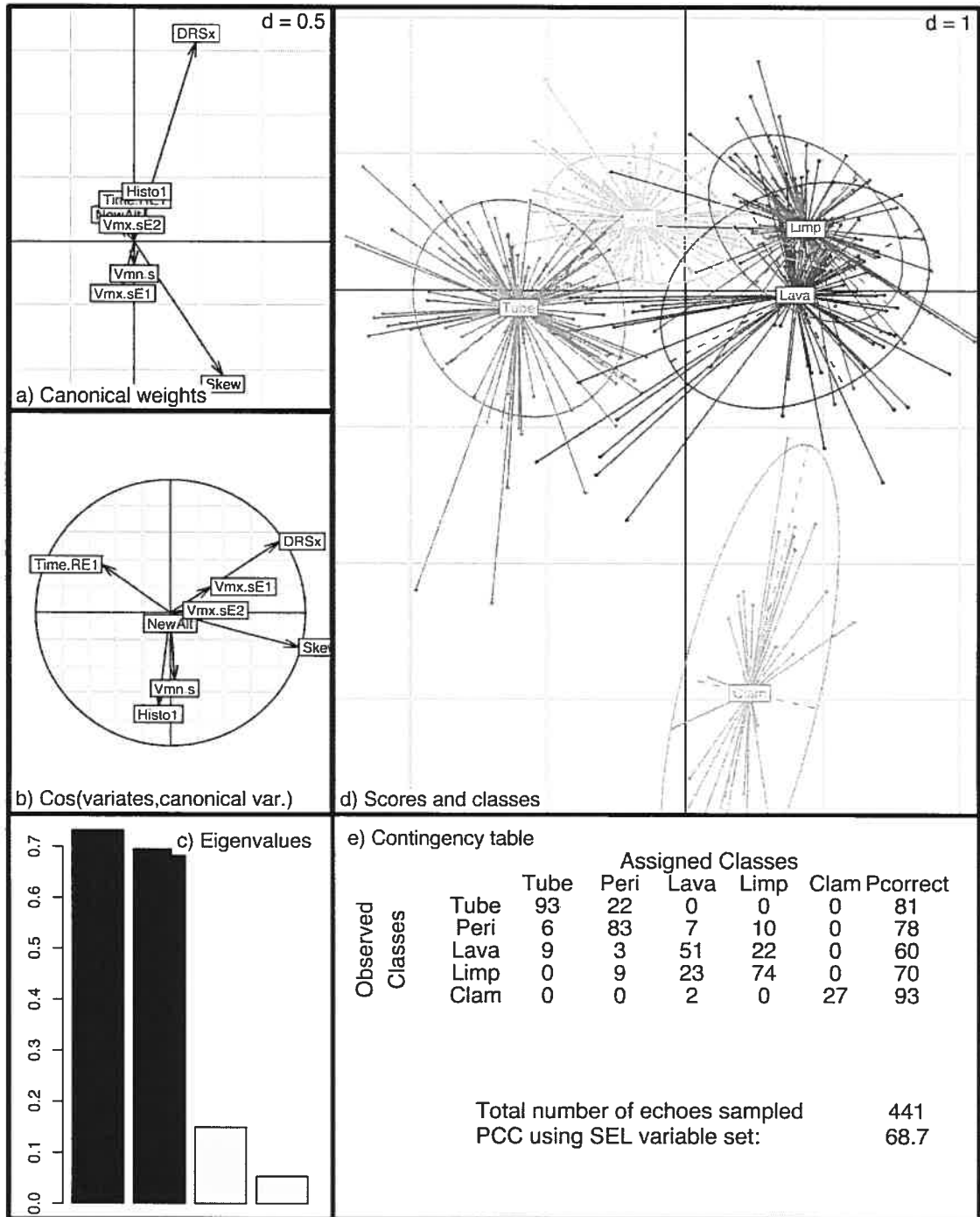


FIG. B.6 – Backscatter discrimination results using VE variables and subtable 7-10 m



---

## APPENDICE C

### **DESCRIPTION AND USAGE OF THE SONAREX PACKAGE WRITTEN IN R LANGUAGE TO ANALYSE THE IMAGENEX BACKSCATTERS.**

---

This beta version of Sonarex package was created to specifically analyse the natures of the echoes found within Imagenex backscatters. Sonarex is an R package executable under the R language, version 2.1.1.

Among the 76 functions created during this Master project which were all used to construct the Sonarex package, we described here only 8 key functions. For more detailed information please load Sonarex package in R console and consult the help menus or the code itself.

---

## C.1 sonarex

*The extraction and analysis of sonar signals for benthic habitat detection purposes*

---

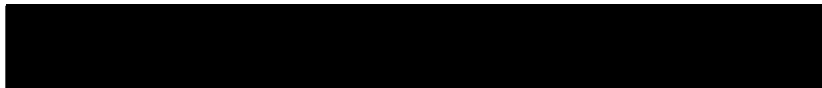
### *Description*

This package is used to filter backscatters and normalize them for depth. Following that, the first and second echoes are found in each backscatter. Three sets of parameters are created from the echo areas. The first data set contains intensity variables, "VI". The second data set contains extracted variables, "VE". The third data set, "VT", contains the union of the VI and VE data sets. See `analyse`. Each variable found in the "VI" data set corresponds to a specific time interval in a backscatter; the variables form a time series describing the echo intensities as the signal rises and falls. The variables in the "VE" dataset are of different types; they describe characteristics of the echo shapes.

HOW TO section :

1. Import and convert the original sonar data into a known and usable format : see `import.imatxt`.
2. Use the `extract` function to perform filtering, correction and variable extraction.
3. Use the `analyse` function to analyse the datasets found in the `extract` function output list.
4. Interpret the results visually using either the `graph.ldaTest` or the `graph.rdaTest` functions.

### *Author(s)*



### *References*

- Clarke, P. A. and L. J. Hamilton. 1999. The ABCS program for the analysis of echo sounder returns for acoustic bottom classification. DSTO-GD-0215, Defence Science & Technology Organisation, Australia.
- Hamilton, L. J. 2001. Acoustic seabed classification systems. Defense Science & Technology Organisation.

Legendre, P. and L. Legendre. 1998. Numerical Ecology. Second English edition. Elsevier, Amsterdam.

Legendre, P., K. E. Ellingsen, E. Bjørnbom and P. Casgrain. 2002. Acoustic seabed classification : improved statistical method. Canadian Journal of Fisheries and Aquatic Sciences 59 : 1085-1089.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. 281-297 in : Le Cam L.M. and J. Neyman [eds.] Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume 1. University of California Press, Berkeley.

### *See Also*

`import.imatxt`, `extract`, `analyse`, `rdaTest`, `ldaTest`, `graph.rdaTest`, `graph.ldaTest`.

---

## C.2 import.imatxt

*This function imports and reads .TXT Imagenex sonar files*

---

### *Description*

This function can import one or several Imagenex(TM) sonar files, so that they can be forwarded to the `extract` function, and later analysed with the `analyse` function.

### *Usage*

```
import.imatxt(  
    path = NA, number.file = NA, echolength = NA, limits = NA, refdate = NA,  
    dateformat = "%d-%b-%Y", timesep = ":", remove.all = TRUE, see = TRUE  
)
```

### *Arguments*

- path** A vector of paths to the .TXT imagenex files. These files were obtained by converting the .81b format files to .TXT. The conversion program used in our case is called 81b2txt.exe and was provided by Imagenex corp. To convert the sonar files to .TXT, users should contact their sonar company to obtain an adapted file converting program.
- number.file** If no path is provided in "path", this parameter defines how many files are to be selected using a browser. The default is NA. If path is defined, its length redefines number.file.
- echolength** How long is each echo? In other words, how many intensity values make up each echo? The default is NA.
- limits** The complete path to a file describing the limits of the data sections that are to be retained for analysis, i.e. their starting and ending times. The file `data(keep.sequences)` shows how such a file should be written. The purpose of these limits is to prevent averaging consecutive echoes acquired in different areas. If the file paths to sonar

transect provided in `path` are individual transects, no value is required here since the time limits of each transect are known; they will be used as limits. The default value is `NA`.

`dateformat`

Date format used. See `strptime` to find out how to use this parameter. The default is `"%d-%b-%Y"`, which stands for day-month abbreviation, and year in four digits.

`timesep`

Time format separator used. Default is `" :"`.

`remove.all`

To make sure all echoes provided are temporally ordered, a filter is used to find all echoes labelled by erroneous or non chronological times. If set to `TRUE`, these echoes are removed; if set to `FALSE`, the erroneous times are shown to the user who is asked to decide what to do. If, after five passes, errors are still found, a warning message appears: `transect were probably listed in the wrong order`.

`see`

Shows the erroneous times, and in which files the erroneous times were found. The default is `TRUE`.

## Details

The transects given must be entered in chronological order.

## Value

`datum`

The sonar data matrix ready to be send to `extract`.

`Column.1`

: Contains the time in seconds past midnight, to which we add 86400 seconds for every day after the reference date.

`Column.2`

: The sonar head angle. 0 degree means facing straight down, perpendicular to the substratum.

`Column.3`

: The estimated acquisition altitude in metres.

`Col4-...`

: The equally-spaced intensity values used to describe the backscatters.

`limits`

The starting and ending times of each separate transect found in the data matrix `"datum"`, in seconds past midnight since `"j.refdate"`. These limits will prevent the averaging of backscatters acquired in different transects or at different locations.

`j.refdate`

A numerical value describing the reference date, which is the earliest date found in the input data sets. This date is represented as the

number of days since 1970-01-01, with negative values for earlier dates. See Dates. E.g. : `j.refdate = as.numeric(Sys.Date())`.

`path` The path of the input files entered.

### *Note*

Please contact me if you are using another type of sonar file. Make sure that in your E-mail, you state the sonar type and program version you are using, and attach a small sample of your sonar data files. These files must be editable, and not under any unusual compressed format. Also, provide a quick definition for all fields (columns) in the file.

### *Author(s)*

[REDACTED]

### *See Also*

extract, analyse, Date

### *Examples*

```
# Usually, the user has first to import their txt files.
# Here is an example, unfortunatly the original files
# could not be stored in this package
# The sonar imagenex file paths are given to R
# path=c("LAVA1", "LAVA2", "LAVA3", "LAVA4", "PERI1", "PERI2",
# "PERI3", "TUBE1", "TUBE2", "#TUBE3", "TUBE4", "LIMP1", "LIMP2",
# "LIMP3", "CLAM1", "CLAM2", "CLAM3", "CLAM4")
# path=paste("~/../ALT", path, ".TXT", sep="")
# The files were imported using import.imatxt function
# vertical.transects<-import.imatxt(path=path, number.file=NA,
# echolength=2000, limits=NA, dateformat="%d-%b-%Y",
# timesep=":", remove.all=TRUE, see=TRUE)
# Creating a smaller subset
# tmp<-sample(1:dim(vertical.transects$datum)[1], 1000)
# vertical.transects$datum<-vertical.transects$datum[tmp[order(tmp)],]
# save(vertical.transects, file="~/../sonarex/data/vertical.transects.rda",
# compress=TRUE)
# That is the data available
```

---

## C.3 extract

*Extraction function that also filters and corrects the sonar data*

---

### *Description*

Prior to the extraction of the two datasets, VE and VI, the data in each backscatter are corrected, filtered, and depth-normalized by this function. "VI".

### *Usage*

```
extract(
  import.datum = NA, save.data=FALSE, save.plot = FALSE, PathOut = NA,
  Prefix = NA, show.echoes = FALSE, show.profiles = FALSE, avegroup = 1.
  aveseq = FALSE, time.filter = NA, timelaps = 13.333, maximumv = 14000.
  pulse = 100, tooweak = 1000, newspacing = 10, reson = 25, soundspeed = 1500,
  alt.trigger = 0.8, depth.ref = 1, nb = 5, dateformat = "%d-%b-%Y", timesep =
  ":", perc.jump = 0.001, thresh.high = 0.05, thresh.low = 0.005
)
```

### *Arguments*

- |                           |   |
|---------------------------|---|
| <code>import.datum</code> | The output list produced by the <code>import.imatxt</code> function.  |
| <code>save.data</code>    | Will the data produced by this function be save in a file, TRUE or FALSE. The default is FALSE, see <code>PathOut</code> .  |
| <code>save.plot</code>    | If <code>save.plot</code> is set to TRUE and <code>show.profiles</code> is also TRUE, a .PDF version of this plot will be created and the plot will not show in the graphics window. The default is FALSE; see <code>PathOut</code> .   |
| <code>PathOut</code>      | The path to the folder where the output files will be created. If <code>Prefix</code> is set to TRUE, the output files will be stored in the “_Results” folder and the basename given in the path will serve as the file prefix. If no path is provided, the “_Results” folder will be created in the folder containing the first file read by the <code>import.imatxt</code> function. |
| <code>Prefix</code>       | A logical value. If <code>Prefix</code> is set to TRUE, the basename given in <code>PathOut</code> will added at the beginning of the name of each output file. If set to FALSE no prefix will be added.  |

**show.echoes**

Produces a graph showing where the program splits and finds the first and second echo starting and ending points of each backscatter. The default is FALSE.

**show.profiles**

Produce a graph showing the depth normalisation effect on the data intensities. The default is TRUE. Because of the large number of points drawn, that plot may take long to produce (the X11 device was found to be much faster at drawing this plot). This plot is very useful to try values and find the proper absorption rate that should be used in order to properly remove the effect of depth embedded in the recorded signals.

**avegroup**

The number of consecutive backscatters to average out. The default is 1 : no averaging is done. If set to 3, each set of three consecutive echoes is averaged.

**avesec**

When set to TRUE, all backscatters found within a one second interval are averaged. The default is FALSE. This setting is used specially when a comparison will be made with a visual survey acquired with one second intervals.

**time.filter**

The complete path of a file containing erroneous time ranges, describing sections that have to be removed from the dataset. The default is NA. This .TXT file contains 4 columns. In the first two columns, we have the date and starting time, while in the third and fourth columns are the ending date and time of the erroneous sections.

**timelaps**

The timelapse between recorded echo intensities. The default is 13.333 microseconds.

**maximumv**

The maximum intensity value that can be recorded by the sonar software. The default is 14000. All intensity values will be converted to the 0 - maximumv intensity scale.

**pulse**

This is the sonar pulse length in microseconds. The default is 100 microseconds.

**tooweak**

While searching for the correct altitude of acquisition value for each backscatter, 'tooweak' defines the lower searching limit. The default is 1000 for a 14000 intensity maximum. Echoes with intensity peaks lower than tooweak will be discarded.



- newspacing** Due to the depth normalisation procedure, time compression or extension is necessary, leading to a necessary intensity resampling. **newspacing** is the new timelapse that will be used to resample the backscatter intensities. The default is 10 microseconds.
- reson** The number of intensity values required to insure that the resonance effect has dissipated from the backscatters. The default is 25; this setting would exclude from the analysis the first 25 intensity values found in every backscatters.
- soundspeed** The speed of sound in water. The default is 1500 metres per second. Set your own value using the following website  
<http://freespace.virgin.net/sd.richards/speed.html>.
- alt.trigger** In each backscatter, once the maximum has been found, a search is initiated to find the first point reaching 80% of that maximum. The altitude of the sonar head will be estimated from a point around this value. The default for **alt.trigger** is 0.8.
- depth.ref** Since the backscatters are depth-normalised, **depth.ref** sets the reference depth. The default is 1 metre.
- nb** This is a search criterion. The program will look for 'nb' consecutively increasing values to determine the 'echo start' point, and 'nb' consecutively decreasing value to determine the 'echo end' point. If no such sequences are found with 'nb', 'nb'-1 will be used, and so forth until a echo starting and ending locations are identified.
- dateformat** This is the date format; see **format.Date** and **strptime** for details.
- timesep** This defines the time format separator. Default is ' : '.
- perc.jump** The percentage of the echo maximum that must be jumped to be considered a valid intensity increase or decrease. The default is 0.001. See **nb**.
- thresh.high** To find an echo start or end position, one of the consecutive **nb** values must be higher than the **thresh.high** percentage. The default is 0.05.
- thresh.low** After the echo starting or ending points have been located, the nearest value found below the 'thresh.low' percentage of the echo maximum is used as a reference for echo start and end detection. The default value is 0.005 %.

*Value*

A list called `extract.data` is returned and saved in PathOut.

"VE"	Data matrix describing all echoes using the following 21 extracted variables :
<i>Date</i>	Backscatter date of acquisition (YYYY-mm-dd).
<i>Time</i>	Backscatter time of acquisition.
<i>Altitude</i>	Computed altitude (serves as ID).
<i>Flag</i>	Defines how many echoes are found in every backscatter.
<i>V.NewAlt</i>	Computed altitude (serves as variable).
<i>Vmx.sE1</i>	Maximum value in the smoothed E1.
<i>Vmn.sE1</i>	Minimum value found between smoothed E1 and E2.
<i>Vmx.sE2</i>	Maximum value in the smoothed E2.
<i>DRSx.E1</i>	Time distance between rise and set area centroids of E1.
<i>DRSy.E1</i>	Intensity distance between rise and set area centroids of E1.
<i>Area.RS.E1</i>	Proportion between E1 rise and E1 set area.
<i>Area.RS.E2</i>	Proportion between E2 rise and E2 set area.
<i>Area.R.E1E2</i>	Proportion between E1 rise and E2 rise area.
<i>Skew.E1</i>	Skewness value derived from the third statistical moment.
<i>Histo1</i>	1st histogram class of E1 intensities, based on matrix max.
<i>Histo...</i>	...
<i>Histo7</i>	7th histogram class of E1 intensities, based on matrix max.
"VI"	A data matrix describing all echoes using intensity variables :
<i>Date</i>	Backscatter date of acquisition (YYYY-mm-dd).
<i>Time</i>	Backscatter time of acquisition.
<i>Altitude</i>	Computed altitude.
<i>Flag</i>	How many echoes are found in every backscatter.
<i>Int1-5</i>	Averaged value of the first five intensities.
<i>Int6-10</i>	Averaged value of the 6th to 10th intensities.
<i>Int...</i>	Continue till all intensity values have been averaged.
<code>call</code>	The call used to run the function.

## Note

This is a beta version ; the smoothing algorithm used will be improved in order to increase calculation speed. The functions used to locate the echo starting and ending points will also be modified.

## Author(s)

[REDACTED]

## References

Clarke, P. A. and L. J. Hamilton. 1999. The ABCS Program for the analysis of echo sounder returns for acoustic bottom classification. Defence Science & Technology Organisation.

Hamilton, L. J. 2001. Acoustic seabed classification systems. Defense Science & Technology Organisation.

## See Also

`import.imatxt`, `analyse`

## Examples

```
# Loading import.imatxt function output
```

```
# Example 1
```

```
data(vertical.transects)
```

```
# These are sonar data acquired through vertical transects where each
```

```
# transect were acquired over only one habitat;
```

```
# five different habitats were sampled.
```

```
s<-extract( import.datum=vertical.transects,
```

```
PathOut="~/ALT.TXT", Prefix=TRUE, save.plot=FALSE,
```

```
avegroup=1, avesec=FALSE, time.filter=NA, timelaps=13.333,
```

```
maximumv=14000, pulse=100, tooweak=1000, newspacing=10,
```

```
reson=25, show.echoes=FALSE, show.profiles=TRUE, soundspeed=1500,
```

```
alt.trigger=0.8, depth.ref=1, nb=5,dateformat="%d-%b-%Y", timesep=":")
```

```
#Example 2
```

```
data(horizontal.transects)
```

```
#These are sonar data acquired through horizontal transects representing
```

```
# the whole habitat gradient.  
s<-extract( import.datum=horizontal.transects,  
PathOut="~/INT.TXT", Prefix=TRUE)
```

---

## C.4 analyse

*Analysis of the matrices of sonar variables created by extract.R.*

---

### *Description*

This function analyses the two matrices generated by the `extract` function. It can compute either a canonical redundancy analysis (RDA) or a linear discriminant analysis (LDA). Cross-validation can be performed, different altitude sections can be analysed separately. If no classification vector (`class.vec`) is provided, several partitions will be found by K-means; the partition that corresponds to the maximum value of the Calinsky-Harabasz criterion will be selected.

### *Usage*

```
analyse(  
  extract.data = NA, VIS = NA, alt.sect = NULL, flag = 0, use = "VT", keep =  
  NA, method = "RDA", class.vec = NULL, validation.perc = 0.3, nperm = 999,  
  criterion = "calinski", select.grpmt = TRUE, cluster = NA, min.g = 2, max.g =  
  25, grpmts.plot = TRUE, dateformat = "%d-%b-%Y", timesep = ":", print.res  
  = TRUE  
)
```

### *Arguments*

`extract.data`

This can be either the loaded output list created by the `extract` function, or the path to the "???.extract.RData" file which contains the list. It can also be left to `NULL`; in that case, the user will be asked to choose a file.

`VIS`

This can be either a loaded matrix or a path to a .TXT table. The information found in this table contains the visual analysis of the footprint areas. Continuous monitoring was performed and the information seen was described for each second of the sonar surveys; see `VIS.analysis`. Note : no `VIS` file will be used with the "LDA" method (see `ldaTest`), but it is necessary when using the "RDA" method (see `rdaTest`). If a `VIS` file is required while this parameter is equal to `NA`, a selection browser will pop up.

<code>alt.sect</code>	This parameter is used when the echo variables from different altitude of acquisition are to be analysed separately. A matrix defining the selected altitude ranges in meters ( $n \times 2$ ) must be provided. Column 1 contains the beginning (inclusive) while column 2 contains the end (exclusive) of each altitude section. Each line corresponds to a separate altitude section on which the analysis will be performed. If NULL, the whole altitude range found in the data set will be used in the analysis.
<code>flag</code>	This parameter indicates which flag state should be used in the analysis.
<code>2</code>	Use the first and second echo sections.
<code>1.5</code>	Use both the first echo and second echo rise sections.
<code>1</code>	Use only the first echo sections.
<code>use</code>	Defines which variable set will be used in the analysis.
<code>"VT"</code>	The analysis is run on the combined set of intensity and extracted variables (combines the VI and VE matrices).
<code>"VE"</code>	The analysis is run on the extracted variables only.
<code>"VI"</code>	The analysis is run on the intensity variables only.
<code>keep</code>	This is either a loaded matrix or a path to a .TXT table. This matrix gives the times sections that are to be use in the analysis. The matrix is of dimension ( $n \times 4$ ). In columns 1 and 2, it provides the date and time of each section starting point ; in the columns 3 and 4 are found the date and time of the end of these sections. Make sure that the date format is the same as specified in <code>dateformat</code> .
<code>method</code>	This is either linear discriminant analysis ("LDA", see <code>ldaTest</code> ) or canonical redundancy analysis ("RDA", see <code>rdaTest</code> ). The "RDA" method requires a matrix of response variables and a set of site visual descriptors describing continuously the studied area. Since the <code>k.means</code> parameter is set to TRUE by default, a partition will be produced and shown in the plot using either ellipses or stars. The "LDA" method requires a grouping vector. If no vector is provided, a partition is produced by K-means.
<code>class.vec</code>	A vector of classes can be provided for every backscatter. If none is provided, one will be created by K-means analysis of the sonar data. If more than one altitude section is given, <code>class.vec</code> will be estimated for each submatrix. This option works only for the "LDA" method.

`validation.perc`

If the "LDA" method is used, a model is produced and written to the output object. If the value of this parameter is 1, 100 % of the data will be used to construct and validate the model. If 0.7 is used, 70 % of the data will be used to construct the model and the other random 30 % will be used to validate the model. If 0 % is used, a pop-up browser will appear and will ask the user to load an existing model and run the data on it (option not implemented yet).

`nperm`

The number of permutations used in the tests of significance.

`criterion`

15 different criteria can be used to decided about the best partition of the data. The default is "calinski". See the help file of the `clustIndex` function.

`select.grpmt`

Allows the user to select its own partition if TRUE (default value).

`min.g`

The minimum number of groups in the K-means partition cascade; the default value is 2.

`max.g`

The maximum number of groups in the K-means partition cascade; the default value is 25.

`grpmts.plot`

Shows a plot of the best partition obtained by K-means, see `criterion`.

`dateformat`

This is the date format; see `format.Date` and `strptime` for details.

`timesep`

This defines the time format separator. The default value is " :".

`print.res`

Either TRUE or FALSE. When TRUE, the results are printed in the R window.

## *Details*

This is a beta version.

## *Value*

A list is returned. For each matrix that was analysed, the results of either `ldaTest` or `rdaTest` are stored in that list. Consequently to reach them, type (for instance) `s$results[[1]]`.

## *Author(s)*



## References

Legendre, P. and L. Legendre. 1998. Numerical Ecology. Second English edition. Elsevier, Amsterdam.

Legendre, P., K. E. Ellingsen, E. Bjørnbom and P. Casgrain. 2002. Acoustic seabed classification : improved statistical method. Canadian Journal of Fisheries and Aquatic Sciences 59 : 1085-1089.

MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. 281-297 in : Le Cam L.M. and J. Neyman [eds.] Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Volume 1. University of California Press, Berkeley.

## See Also

```
import.imatxt, extract.rdaTest, ldaTest.
```

## Examples

```
data(ALTextract)

#-----#
# ANALYSIS OF A VERTICAL TRANSECTS #
#-----#

# Constructing a partition vector based on the acquisition times

times<-time2spm(date=extract.data$VE[,1], time=extract.data$VE[,2],
dateformat="%Y-%m-%d")
lava<- time2spm(date=c("19-AUG-2002", "19-AUG-2002"),
time=c("19:55:02.64", "20:23:34.55"))
peri<- time2spm(date=c("19-AUG-2002", "19-AUG-2002"),
time=c("20:23:34.56", "20:52:31.90"))
tube<- time2spm(date=c("19-AUG-2002", "19-AUG-2002"),
time=c("20:52:31.91", "21:32:00.67"))
limp<- time2spm(date=c("19-AUG-2002", "19-AUG-2002"),
time=c("21:32:00.68", "21:54:44.80"))
clam<- time2spm(date=c("19-AUG-2002", "19-AUG-2002"),
time=c("21:54:44.81", "22:22:48.39"))
class.vec<-rep(0,length(times))
for (i in 1:length(times)){
```



```

if(times[i]>=lava[1] & times[i]<lava[2]) class.vec[i]<-1
if (times[i]>=peri[1] & times[i]<peri[2]) class.vec[i]<-2
if (times[i]>=tube[1] & times[i]<tube[2]) class.vec[i]<-3
if (times[i]>=limp[1] & times[i]<limp[2]) class.vec[i]<-4
if (times[i]>=clam[1] & times[i]<clam[2]) class.vec[i]<-5
}
v<-class.vec
v[v==1]<-"Lava";
v[v==2]<-"Peri";
v[v==3]<-"Tube";
v[v==4]<-"Limp";
v[v==5]<-"Clam";
class.vec<- factor(v)
# Linear discriminant analysis (LDA) is used for the analysis
s<-analyse(extract.data=extract.data, VIS=NA,
alt.sect=NULL, use= "VT", keep=NA, class.vec=class.vec,
validation.perc=0.30,
nperm=9, criterion="Calinski", select.grpmt=TRUE, flag=2,
method="LDA", grpmts.plot=TRUE)

#-----#
# ANALYSIS OF A TYPICAL HORIZONTAL TRANSECTS #
#-----#
data(INTextract)
# Along with a visual transect
data(VIS.analysis)
# Redundancy analysis (RDA) is used for the analysis
s<-analyse(extract.data=extract.data, VIS=VIS.analysis,
alt.sect=NULL, use="VT", keep=NA, class.vec=NULL,
validation.perc=0.30, nperm=9,
criterion="Calinski", select.grpmt=FALSE, flag=2,
method="RDA", grpmts.plot=TRUE)

```

---

## C.5 ldaTest

*Compute a linear discriminant analysis*

---

### *Description*

This function performs a linear discriminant analysis and complements the results with some statistical results which facilitate the interpretation.

### *Usage*

```
ldaTest(  
  X.mat, group, print.res = TRUE, alpha = 0.05, cross = NULL  
)
```

### *Arguments*

- |                        |   |
|------------------------|---|
| <code>X.mat</code>     | <code>X.mat</code> (nxm) contains the explanatory variables. The number of variables 'm' is computed after eliminating collinear explanatory variables, if any.   |
| <code>group</code>     | A vector or factor of length (n) describing the group to which each object described in <code>X.mat</code> belongs.   |
| <code>print.res</code> | If set to <code>TRUE</code> , print most of <code>ldaTest</code> output to the screen.  |
| <code>alpha</code>     | Statistical threshold; default is 5%.   |
| <code>cross</code>     | Describes the percentage of the objects that will be used to create the model; cross-validation will be performed using the remaining objects. The default is <code>NULL</code> ; in that case 100% of the objects are used to create the model and 100% are used to test it. |

### *Value*

The function returns an output list containing the following elements :

- |                        |  |
|------------------------|--|
| <code>VIF</code>       | Variance inflation factors for the explanatory variables X; the value 0 designates entirely collinear variables. |
| <code>canEigval</code> | Canonical eigenvalues.   |
| <code>C</code>         | Matrix C (pxk) containing the "Final vectors for the explanatory variables".                                     |

<code>F.model</code>	Matrix ( $n1 \times k$ ) of object scores for the constructed model.
<code>F.predict</code>	( $n \times k$ ) Matrix of object scores for the object used in the prediction.
<code>Stats</code>	
<code>Homo</code>	Using Kullback's statistic, we test the hypothesis of homogeneity of the within-group dispersion matrices.
<code>Axes</code>	Results of Wilks' ratio test of significance for all axes together, then after removing the first axis, and so on until the remaining axes are no longer significant.
<code>Class.Ftn</code>	Matrix ( $k \times p$ ) of weights given by the canonical functions.
<code>Scores</code>	Matrix describing the scores obtained using the classification functions with the objects. It <code>cross</code> is not NULL nor 100%, the <code>Scores</code> matrix describes only the object that were not used to construct the discriminant model.
<code>Grp.Mod</code>	The groups originally observed for the ( $n1$ ) objects used to construct the model.
<code>Grp.Obs</code>	The groups originally observed for the ( $n2$ ) objects used to validate the model.
<code>Grp.Ass</code>	The groups assigned, using the model, to the ( $n2$ ) objects used to validate the model.
<code>Grp.Ctr.Mod</code>	
	Group centroids based on the <code>Grp.Mod</code> groups.
<code>Grp.Ctr.Mod</code>	
	Group centroids based on the <code>Grp.Obs</code> groups.
<code>Grp.Ctr.Mod</code>	
	Group centroids based on the <code>Grp.Ass</code> groups.
<code>Class.Table</code>	
	The group classification (or "confusion") table.

### *Author(s)*

Pierre Legendre, Sebastien Durand, Universite de Montreal.

### *References*

Legendre, P. and L. Legendre. 1998. Numerical Ecology. Second English Edition. Elsevier, Amsterdam.

*See Also*

```
graph.ldaTest
```

*Examples*

```
# Example 1, from Legendre & Legendre (Numerical ecology, 1998) p. 626
XY.mat<-matrix(c(1,2,2,8,8,8,9,2,2,1,7,6,3,3),7,2,byrow=FALSE)
groups<-c(1,1,1,2,2,3,3)
results <- ldaTest(XY.mat,group=groups)
graph.ldaTest(results, ell=FALSE)

# Example 2
XY.mat2<-rbind(XY.mat,XY.mat+0.5,XY.mat-0.5,
XY.mat+0.2,XY.mat-0.2,XY.mat+0.3,XY.mat-0.3)
groups2=c("La","La","La","BE","BE","CO","CO")
groups2=c(rep(groups2,7))

colnames(XY.mat2)=c("El11","El12")
rownames(XY.mat2)=paste("a",1:nrow(XY.mat2))

# Construct a model using 40% of the data and use the remaining 60% for
# prediction by cross-validation
results <- ldaTest(XY.mat2, group=groups2,
print.res=FALSE,cross=0.4,alpha=0.9)

# Plot the model with the observed groups; reverse axis 1
graph.ldaTest(results,plot.what="model", xax=-1, yax=2, lty.ell=2,
col.ell="dark green", ell.axis=TRUE, pos.ctd=4)

# Plot the predictions with the assigned groups.
# Variable El11 in column 1 is binary
graph.ldaTest(results, xax=1, yax=2, binary=1, pos.ctd=1)

# The ellipse drawn seem large. This is because they represent a 95%
# confidence intervals and the number of objects is small.
```

---

## C.6 rdaTest

*Compute a canonical redundancy analysis*

---

### *Description*

This function computes simple redundancy analysis (RDA) for partial RDA, with permutation tests, following the algorithm described in Numerical Ecology, Chapter 11 (Legendre & Legendre, 1998).

### *Usage*

```
rdaTest(  
  YY.mat, XX.mat, which.cov = NULL, scale.Y = FALSE, testF = NULL, nperm  
  = NULL, print.res = TRUE, print.cum = FALSE  
)
```

### *Arguments*

- |                        |   |
|------------------------|---|
| <code>YY.mat</code>    | The (n $\times$ p) site-by-species data table.  |
| <code>XX.mat</code>    | The (n $\times$ m) table of explanatory variables. The number of variables (m) is recomputed after eliminating collinear explanatory variables, if any. Covariables, if any, are also found in that table.            |
| <code>which.cov</code> | A vector (1 $\times$ max(m-1)) defining the covariables by their column numbers in XX.mat.  |
| <code>scale.Y</code>   | A logical value (TRUE or FALSE) defining if YY.mat should be standardized (TRUE), or only centred on the column means (FALSE).  |
| <code>testF</code>     | If NULL, the program will ask the user if he/she wishes to test the F statistic. If testF is TRUE or FALSE, no question will be asked; the program will perform the test, or not, in accordance with that indication. |
| <code>nperm</code>     | Number of permutations for the F test. If NULL, a question will be asked by the program.  |
| <code>print.res</code> | Prints most of the rdaTest output on the screen.  |
| <code>print.cum</code> | Prints the fractions of the response variable's (e.g. species) variances explained by canonical axes 1, 2, 3, ... and by the whole canonical analysis.  |

*Value*

If `print.res = TRUE`, the function prints the following information to the R window :

1. The variance inflation factors for the explanatory variables (the covariables are not included in these calculations).
2. The bimultivariate redundancy statistic (canonical R-square), as well as the adjusted R-square when there are no covariables in the analysis.
3. Test of significance of the canonical relationship : the F statistic and permutational probability. Note : the degrees of freedom of F are not corrected for collinearity between the explanatory variables and the covariables. This has no influence on the associated permutational probability.
4. The number of objects, number of response variables, and number of explanatory variables after removing collinear variables ; the number of canonical eigenvalues larger than 0.
5. The total variance in matrix `YY.mat`, i.e., the  $SS/(n-1)$ .
6. The eigenvalues, relative eigenvalues, and the cumulative % the variance of species data accounted for by the successive canonical eigenvalues.

The function also returns an output list containing the following ELEMENTS :

<code>VIF</code>	Variance inflation factors for the explanatory variables X ; the value is 0 for entirely collinear variables. The covariables are not included in this calculation.
<code>canEigval</code>	Canonical eigenvalues.
<code>U</code>	(pxk) Canonical eigenvectors normalized to 1 (scaling 1).
<code>USc2</code>	(pxk) Canonical eigenvectors normalized to $\sqrt{\text{eigenvalue}}$ (scaling 2).
<code>F</code>	(nxk) Matrix of object scores (scaling 1).
<code>Z</code>	(nxk) Matrix of fitted object scores (scaling 1).
<code>FSc2</code>	Matrix of object scores (scaling 2).
<code>ZSc2</code>	Matrix of fitted object scores (scaling 2).
<code>biplotScores1</code>	Biplot scores of explanatory variables (scaling 1).
<code>biplotScores2</code>	Biplot scores of explanatory variables (scaling 2).
<code>FitSpe</code>	Table of cumulative fit per species (in %) as fraction of variance of species.

VarExpl	Vector of total % fit per species after all canonical axes.
ProbFrda	Probability associated with F test of the canonical relationship.
X.mat	Original X matrix (required by the plotting function).
AxisVar	Eigenvalues as in Canoco : fraction of variance of YY.mat explained by each canonical axis.

### *Author(s)*

Pierre Legendre and Sebastien Durand, Universite de Montreal.

### *References*

Legendre, P. and L. Legendre. 1998. Numerical Ecology. Second English Edition. Elsevier, Amsterdam.

### *See Also*

`graph.rdaTest`

### *Examples*

```
# Example from Legendre & Legendre (1998), p. 590, Table 11.3
data(coral)
Table = as.matrix(coral)
Y=Table[,1:6]
X=Table[,7:10]

# Numerical Ecology
resultats <- rdaTest(Y,X,nperm=99,testF=TRUE)
graph.rdaTest(resultats,plot.type="F",stars=FALSE,lty.ell=3, centroid=FALSE)

#or
graph.rdaTest(resultats,plot.type="F",lty.ell=3,pos.site=4, mai.perc=0.15)

resultats <- rdaTest(Y,X,testF=TRUE,nperm=9, scale.Y=FALSE,
print.res=FALSE, print.cum=FALSE)
graph.rdaTest(resultats,xax=-1,yax=2, mul.spc=0.90,
mul.env=0.70, mul.text=0.10, scaling=1,
plot.type="F",mai.perc=0.15, stars=FALSE, pos.site=4, centroid=FALSE)
```

---

## C.7 graph.ldaTest

*Display ldaTest results*

---

### *Description*

This is a highly flexible drawing device for discriminant analysis results.

### *Usage*

```
graph.ldaTest(  
  ldaTest, xax = 1, yax = 2, plot.what = "p.assign", binary = NULL, centroids  
  = TRUE, ell = TRUE, sites = TRUE, stars = TRUE, alpha = 0.95, filter.env  
  = NULL, saveplot = FALSE, path = NULL, width = 6, height = 6, xlim =  
  NULL, ylim = NULL, lwd = 1, len = 0.1, lty.env = 1, lty.ell = 1, lty.axis = 2,  
  lty.stars = 1, col.env = "blue", col.group = NULL, col.ell = "black", col.ctd =  
  "black", pos.env = 4, pos.bin = 4, pos.site = NULL, pos.ctd = 4, cex.env = 1,  
  cex.site = 1, cex.ctd = 1, cex.axis = 1, pch.bin = 21, pch.site = 21, pch.ctd =  
  23, ell.axis = FALSE, mul.env = 0.9  
)
```

### *Arguments*

- |                        |  |
|------------------------|--|
| <code>ldaTest</code>   | The <code>ldaTest</code> output.   |
| <code>xax</code>       | The matrix column to be used as the abscissa of the lda graph. This value can be negative; if so, the plot axis will be reversed.  |
| <code>yax</code>       | The matrix column to be used as the ordinate of the lda graph. This value can be negative; if so, the plot axis will be reversed.  |
| <code>plot.what</code> | Can be set to either "model", "observe", or "assign". If "model" is selected, the data used to construct the model as well as the original (observed) groups are shown. If "observe" is used, the data used for validation are plotted along with the observed groups. Finally and the default, the data used in model validation are plotted along with the groups assigned to them by the model. |
| <code>binary</code>    | A vector specifying which of the environmental variables are binary. The data points possessing the characteristics in question will be averaged and represented by a symbol instead of an arrow, as in Canoco.  |



<code>centroids</code>	If set to TRUE, group centroids are drawn.
<code>ell</code>	Draw confidence ellipses around groups of points defined by a vector defining the group assignments of the objects; see <code>alpha</code> below.
<code>sites</code>	Shows by default the sites on the plot; "FALSE" will not display the sites.
<code>stars</code>	Shows by default stars on the plot; stars are lines linking each site to its respective group centroid. "FALSE" will not display them.
<code>alpha</code>	The confidence region of the ellipses (e.g. : 0.90, 0.80, etc.) Default value : 0.95.
<code>filter.env</code>	A percentage value which filters out environmental variables associated to very short vectors. Based on the total length of all environmental variables, only those whose lengths are greater than the specified percentage will be displayed. The default is NULL : no filtering.
<code>saveplot</code>	The plot can be saved to a .pdf file. If set to FALSE, the plot will only be plotted in a window. If set to TRUE, no graph will be plotted on the screen, but a pdf file will be created.
<code>path</code>	Complete path to the file in which the plot will be saved; example : “ /Desktop/toto.pdf”.
<code>width</code>	Plot width in inches.
<code>height</code>	Plot height in inches.
<code>xlim</code>	Vectors describing the x axis minimum and maximum of the plotted region.
<code>ylim</code>	Vectors describing the y axis minimum and maximum of the plotted region.
<code>lwd</code>	Line width for the axes, arrows, and ellipses.
<code>len</code>	Length of the arrow heads.
<code>lty.env</code>	The environmental variables arrow drawing type.
<code>lty.spc</code>	Species variables arrow drawing type.
<code>lty.ell</code>	The drawing type for ellipses; see 'lty' in the help documentation file of <code>par</code> .
<code>lty.axis</code>	The drawing type of axes, dotted lines, etc. See <code>par</code> .
<code>lty.stars</code>	The line type for stars. See <code>par</code> .
<code>col.env</code>	The environmental variables arrow and text color.

<code>col.group</code>	By default, each group is assigned a different colour, so that sites and group centroids have an identifiable colour, which is a function of their group membership.
<code>col.ell</code>	Colour of the ellipses.
<code>col.ctd</code>	Colour of the centroids.
<code>pos.env</code>	Position offset for the names of the environmental variables : 1 = bottom, 2 = left, 3 = top, 4 = right.
<code>pos.bin</code>	Position offset of the binary environmental variables names : 1=bottom, 2=left, 3=top, 4=right.
<code>pos.site</code>	Position offset of the site names : 1=bottom, 2=left, 3=top, 4=right.
<code>pos.ctd</code>	Position offset of the group centroid names : 1=bottom, 2=left, 3=top, 4=right.
<code>cex.env</code>	Font size for environmental variable labels.
<code>cex.site</code>	Font size for sites labels.
<code>cex.ctd</code>	Font size for centroid labels.
<code>cex.axis</code>	Font size for axis names.
<code>pch.bin</code>	Symbol used to represent binary variables. See <code>pch</code> definition in the <code>points</code> function.
<code>pch.site</code>	Symbol used to represent sites. See <code>pch</code> definition in the <code>points</code> function.
<code>pch.ctd</code>	Symbol used to represent centroids. See <code>pch</code> definition in the <code>points</code> function.
<code>ell.axis</code>	If TRUE, draw the major and minor axes of the ellipses.
<code>mul.env</code>	The environmental variable arrows are initially expanded or compressed to fit within the plot borders and be displayed with maximum spread. <code>mul.env</code> is a multiplier used to modify the arrow proportions. By default, 0.90 is used to give some room for the arrow labels.

### *Value*

A plot is returned.

### *Author(s)*



*See Also*

ldaTest

*Examples*

```
data(lda.analyse)
# Default
graph.ldaTest(s$results[[1]])

# Without the ellipses and sites
graph.ldaTest(s$results[[1]], ell=FALSE, sites=FALSE, cex.ctd=2)

# Without in gray and without the stars
graph.ldaTest(s$results[[1]], pch.site=10, col.group=gray(0:3/3),
stars=FALSE)

# With binary variables the ten first, and some visual filtering of weak
# environmental variables

graph.ldaTest(s$results[[1]], col.group=gray(0:3/3), stars=FALSE,
ell.axis=TRUE, col.ctd="red", binary=c(1:10), filter.env=0.04,
cex.ctd=1.5)

# In a zoomed window
graph.ldaTest(s$results[[1]], sites=FALSE, col.ctd="red",
xlim=c(-2,2), ylim=c(-2,2), pos.env=3)

# Having filtered the environmental variables that does not show
# much influence on the displayed axis
graph.ldaTest(s$results[[1]], col.group=gray(0:3/3), stars=FALSE,
sites=FALSE, ell.axis=FALSE, col.ctd="red", filter.env=0.05, pos.bin=4,
pos.env=2, lty.axis=0)
```

---

## C.8 graph.rdaTest

*Display the rdaTest output*

---

### *Description*

This is highly flexible drawing device

### *Usage*

```
graph.rdaTest(
  rdaTest, xax = 1, yax = 2, scaling = 1, plot.type = "notchosen", binary =
  NULL, groups = NULL, centroids = TRUE, ell = TRUE, sites = TRUE, stars
  = FALSE, label.env = TRUE, label.spc = TRUE, label.site = TRUE, alpha =
  0.95, saveplot = FALSE, path = NULL, width = 6, height = 6, xlim = NULL,
  ylim = NULL, lwd = 1, len = 0.1, lty.env = 1, lty.spc = 1, lty.ell = 1, lty.stars
  = 1, lty.axis = 2, col.env = "blue", col.spc = "red", col.groups = NULL, col.ell
  = "black", col.ctd = "black", pos.env = NULL, pos.bin = 4, pos.site = NULL,
  pos.ctd = 4, cex.env = 1, cex.spc = 1, cex.site = 1, cex.ctd = 1, cex.axis = 1,
  pch.bin = 21, pch.site = 21, pch.ctd = 23, ell.axis = FALSE, mul.spc = 0.90,
  mul.env = 0.90, mul.text = 0.10, mai.perc = NULL
)
```

### *Arguments*

<code>rdaTest</code>	The <code>rdaTest</code> output.
<code>xax &amp; yax</code>	The matrix columns to be used as the abscissa and ordinate in the rda graph. These value can be negative ; if so, the plot axes will be reversed.
<code>scaling</code>	Allows the user to choose between the two scaling types, 1 or 2.
<code>plot.type</code>	Either "F" = site scores, or "Z" = fitted site scores.
<code>binary</code>	A vector specifying which of the environmental variables are binary. The data points possessing the characteristics in question will be averaged and represented by a symbol instead of an arrow, as in Canoco.
<code>groups</code>	A vector describing the group assignment of all objects.

centroid	A logical value (TRUE or FALSE) defining if centroids are to be drawn.
ell	A logical value (TRUE or FALSE) defining if confidence ellipses around groups of points are to be drawn. Group membership is described in "groups". See examples below.
sites	A logical value (TRUE or FALSE) defining if the sites are to be drawn on the plot.
stars	A logical value (TRUE or FALSE) defining if stars (lines linking each site to its respective group centroid) are to be drawn on the plot.
label.env	A logical value (TRUE or FALSE) defining if the environmental variable names are to be printed on the plot.
label.spc	A logical value (TRUE or FALSE) defining if the species names are to be printed on the plot.
label.site	A logical value (TRUE or FALSE) defining if the site names are to be printed on the plot.
alpha	The confidence region of the ellipses (e.g. : 0.90, 0.80, etc.) Default value : 0.95.
saveplot	The plot can be saved to a .pdf file. If set to FALSE, the plot will only be plotted in a window. If set to TRUE, no graph will be plotted on the screen, but a pdf file will be created.
path	Complete path to the file in which the plot will be saved ; example : “ /Desktop/toto.pdf”.
width & height	Plot width and height in inches.
xlim & ylim	Vectors describing the x and y axes minimum and maximum of the plotted region.
lwd	Line width for the axes, arrows, and ellipses.
len	Length of the arrow heads.
lty.env	The environmental variables arrow drawing type.
lty.spc	The species variables arrow drawing type.
lty.ell	The drawing type for ellipses ; see 'lty' in the help documentation file of 'par'.
lty.stars	The line type for stars. See par.

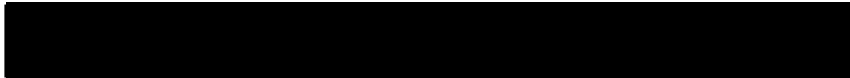
lty.axis	The drawing type of axes, dotted lines, etc. See par.
col.env	The environmental variables arrow and text colour.
col.spc	The species variables arrow and text colour.
col.group	By default, each group is assigned a different colour, so that sites and group centroids have an identifiable colour, which is a function of their group membership.
col.ell	Colour of the ellipses.
col.ctd	Colour of the centroids.
pos.env	Position offset of the environmental variable names : 1=bottom, 2=left, 3=top, 4=right.
pos.bin	Position offset of the binary environmental variables names : 1=bottom, 2=left, 3=top, 4=right.
pos.site	Position offset of the site names : 1=bottom, 2=left, 3=top, 4=right.
pos.ctd	Position offset of the group centroid names : 1=bottom, 2=left, 3=top, 4=right.
cex.env	Font size for environmental variable labels.
cex.spc	Font size for species names.
cex.site	Font size for sites labels.
cex.axis	Font size for axis names.
pch.bin	Symbol used to represent binary variables. See pch definition in the points function.
pch.site	Symbol used to represent sites. See pch definition in the points function.
pch.ctd	Symbol used to represent centroids. See pch definition in the points function.
ell.axis	If TRUE, draw the major and minor axes of the ellipses.
mul.spc	The species variable arrows are initially expanded or compressed to fit within the plot borders and be displayed with maximum spread. mul.spc is a multiplier used to modify the arrow proportions. By default, 0.90 is used to give some room for the arrow labels.
mul.env	The same as mul.spc but for the environmental variable.
mul.text	Text position at the end of each arrow, plus a percentage of the arrow length; the default value is 10%.

`mai.perc` Allows to increase or decrease the margin size, using a percentage of the total width and height of the graph. That value is between 0 and 1.

### *Value*

A plot is returned.

### *Author(s)*



### *See Also*

`rdaTest`

### *Examples*

```
#see \code{\link{rdaTest}}
data(rda.analyse)
graph.rdaTest(s$results[[1]], plot.type = "Z")

#Find groups
toto<-wrapkmeans(s$results[[1]][[6]],2,5,iter=9,criterion="calinski")
toto2<-plotkmeans(toto)
#in this case we take the 5 group solution
graph.rdaTest(s$results[[1]], plot.type = "Z", groups=toto[[1]][,4],
label.site=FALSE)

graph.rdaTest(s$results[[1]], plot.type = "Z", groups=toto[[1]][,4],
lty.env=0, lty.spc=0, ell=FALSE, stars=TRUE, label.site=FALSE,
xlim=c(-1,1), ylim=c(-1,1))
```

