



HAL
open science

Analyse et évaluation de la fiabilité de l'information dans le système d'information géographique des assemblages benthiques méditerranéens "MBA"

Axel Iehlé, Lucien Wald, Charles-François Boudouresque

► To cite this version:

Axel Iehlé, Lucien Wald, Charles-François Boudouresque. Analyse et évaluation de la fiabilité de l'information dans le système d'information géographique des assemblages benthiques méditerranéens "MBA". Scientific Reports of the Port-Cros National Park, 1995, 16, pp.93-113. hal-00464748

HAL Id: hal-00464748

<https://minesparis-psl.hal.science/hal-00464748>

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ANALYSE ET EVALUATION DE LA FIABILITE DE L'INFORMATION DANS LE SYSTEME D'INFORMATION GEOGRAPHIQUE DES ASSEMBLAGES BENTHIQUES MEDITERRANEENS «MBA»

IEHLE A.^a , WALD L.^a , BOUDOURESQUE C.-F.^b

Résumé : La présente étude concerne le Système d'Information Géographique «MBA» sur les assemblages benthiques marins de Méditerranée. Ce système contient sous une forme numérique les cartes publiées sur la région de La Ciotat (Bouches-du-Rhône, France) à l'île du Levant (Var, France) et permet de faire la synthèse de ces informations redondantes et conflictuelles. La carte synthétique de l'île de Port-Cros (Var, France) est présentée. Deux grands types d'erreurs, dans la construction de la base de données et dans la synthèse de l'information, sont définis, à l'intérieur desquels chaque source potentielle d'imperfection de l'information est décrite et une solution est proposée. Une nouvelle technique est présentée permettant l'automatisation de l'évaluation de la fiabilité des documents utilisés. Pour cela, on réalise une classification de l'image en fonction de cartes de paramètres influant sur la répartition des assemblages benthiques.

Abstract : The present study deals with the Geographic Information System «MBA» for benthic communities of the Mediterranean sea. This system contents, on a digital form, the maps published concerning the region from La Ciotat (Bouches-du-Rhône, France) to the Levant island (Var, France) and allows to make a synthesis of these redundant and conflicting informations. The synthetic map of the Port-Cros island (Var, France) is presented. Two main types of error, in the construction of the database and during the synthesis of the information, are defined in which each potential source of imperfection of the information is described and discussed. A new technique is presented allowing automatation of the evaluation of the fiability of the documents used. For that, a classification of the image is made using parameters having an influence upon the distribution of the benthic communities.

INTRODUCTION

La cartographie des peuplements benthiques marins est le point de départ de nombreuses études en écologie, sédimentologie, et aménagement du littoral. Les techniques utilisées en Méditerranée

^a : Centre de Télédétection et d'Analyse des Milieux Naturels, Ecole Nationale Supérieure des Mines de Paris, Sophia-Antipolis, 06565 Valbonne Cedex, France.

^b : LBMEB (Laboratoire de Biologie Marine et d'Ecologie du Benthos), Faculté des Sciences de Luminy, 13288 Marseille Cedex 9, France.

comprennent essentiellement l'interprétation de photographies aériennes, les échantillonnages par drague ou benne, ou encore l'observation directe en scaphandre autonome (MEINESZ *et al.*, 1981 ; COLANTONI *et al.*, 1982 ; BOUDOURESQUE *et al.*, 1985 ; RAMOS-ESPLA, 1985). Il en résulte des cartes de finalités, d'échelles et de fiabilité variées, qui sont difficilement comparables. Les synthèses cartographiques manuelles jusqu'à présent disponibles (JEUDY DE GRISSAC *et al.*, 1985 ; PERGENT et PERGENT, 1988) sont fastidieuses à réaliser, et difficiles à critiquer car les processus de synthèse propres à l'auteur sont rarement expliqués.

En 1986 a été mis en place au Centre de Télédétection et d'Analyse du Milieu Naturel (CTAMN) de l'Ecole Nationale Supérieure des Mines de Paris (ENSM) à Sophia Antipolis, en collaboration avec le Laboratoire de Biologie Marine et d'Ecologie du Benthos (LMBEB) de la Faculté des Sciences de Marseille-Luminy, un Système d'Information Géographique (SIG) axé sur l'étude des assemblages benthiques méditerranéens.

Un SIG peut être considéré comme une base de données dont la plupart sont repérées géographiquement et qui est associée à un système interactif de gestion et d'analyse de ces données. L'utilisation de l'outil informatique en cartographie permet d'obtenir des documents standardisés pour ce qui est de la géométrie, de l'échelle et des codes représentant l'information contenue par les cartes stockées.

Cet article fait suite aux travaux de MEAILLE *et al.* (1988) qui décrivent les étapes de la constitution de la base de données du SIG «MBA», et de MEAILLE (1988) présentant la méthode actuelle de synthèse cartographique. Hormis l'augmentation de la surface cartographiée, notre étude porte essentiellement sur la question de la fiabilité des informations stockées ou produites. En effet, la première utilisation qui est faite de ce SIG est la production d'une carte de synthèse dont la qualité est très dépendante de celle des documents de base utilisés. Nous avons donc analysé les sources d'imperfections à chaque étape de la constitution de la carte synthétique. Ces erreurs ont été quantifiées lorsque cela était possible et des méthodes pour les diminuer ou les annuler, proposées.

MATERIEL ET METHODES

Le système installé au CTAMN est organisé autour d'un micro-ordinateur de type IBM-compatible et a été décrit par MEAILLE *et al.* (1988). Le stockage de l'information s'effectue par l'intermédiaire d'une table à numériser sur disque optique numérique d'une capacité de 100 méga-octets. L'affichage des cartes numériques se fait sur un écran

graphique couleur, et une imprimante couleur permet d'en obtenir une copie papier.

La zone étudiée est un rectangle de 23 km par 82 km (Fig. 1) dont les coordonnées géographiques dans le repère Lambert zone III utilisé sont 3105N à 3082N et 861E à 943E, soit 43°12' N à 42°58' N et 5°33' E à 6°33' E.

Cette surface a été découpée en treize secteurs afin d'adapter les cartes aux possibilités graphiques des micro-ordinateurs en diminuant la proportion du domaine terrestre sur la surface totale étudiée, économisant ainsi la place nécessaire au stockage. Elle s'étend de La Ciotat (Bouches-du-Rhône, France) à l'île du Levant (Var, France).

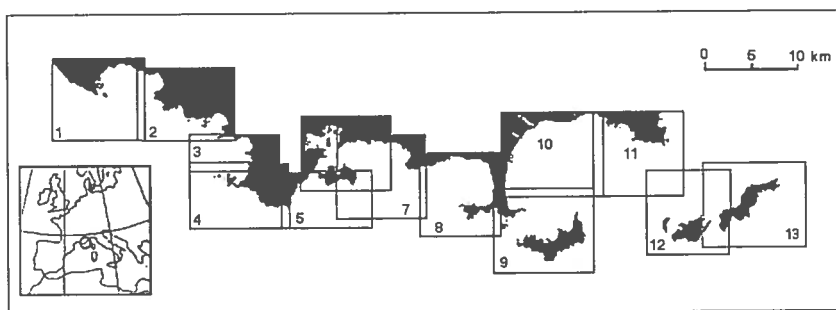


Fig. 1 : Vue générale de la surface couverte par la base de données, séparée en secteurs (numérotés de 1 à 13).

Pour chaque secteur analysé, les cartes de peuplements benthiques publiées ont d'abord été sélectionnées pour éliminer les documents redondants ou synthétiques, trop vieux ou d'échelle inappropriée, puis l'information a été codée de façon standard en vingt-trois classes rassemblées en dix groupes (Tableau I).

Chaque carte est alors numérisée dans sa géométrie d'origine puis rectifiée géométriquement pour s'ajuster au contour de côte IGN 1/25 000ème choisi comme référence.

La banque de données présente, à ce stade, une structure multi-couche constituée par l'empilement des cartes numérisées en autant de canaux (Fig. 2). Cette organisation un peu spéciale, car présentant une redondance de l'information entre les canaux d'un même secteur, est imposée par l'impossibilité d'avoir une vue directe d'ensemble du benthos. L'incertitude vis-à-vis de la réalité des données est compensée par la surimposition de plusieurs «avis» sur la répartition des assemblages benthiques sur ce secteur. La numérisation d'une nouvelle carte engendre une amélioration de la qualité (ou tout au moins une augmentation des chances de qualité) de l'ensemble.

Tableau I — Table de codification des assemblages benthiques

Code	Classe
0 : NC	Non cartographié
1 : AP	Algues photophiles
2 : AP/d	Algues photophiles dégradées
3 : AP/e	Algues photophiles sur éboulis sous-marins
10 : P	Milieu portuaire ou fortement pollué
11 : SVMC	Sables vaseux de mode calme
12 : SVMC/z	Herbier à <i>Zostera noltii</i> sur SVMC
13 : SVMC/c	Herbier à <i>Cymodocea nodosa</i> sur SVMC
20 : SFHN	Sables fins de hauts niveaux
21 : SFBC	Sables fins bien calibrés
22 : SGCF	Sables grossiers et fins graviers sous influence de courants de fond
30 : HP	Herbiers à <i>Posidonia oceanica</i>
31 : HP/d	Herbiers à <i>Posidonia oceanica</i> dégradés
32 : MM	Mattes mortes de <i>Posidonia oceanica</i>
40 : C	Fonds coralligènes (fortement bioconcrétionnés)
41 : PC	Précoralligènes (faiblement bioconcrétionnés)
50 : DC	Détritique côtier
51 : DC/c	Détritique côtier plus ou moins bioconcrétionné
52 : DC/p	Détritique côtier à Peyssonneliacées libres
53 : DC/m	Détritique côtier avec maërl ou pralines
54 : DC/i	Détritique côtier avec faciès d'instabilité
60 : DE	Détritique envasé
70 : VTC	Vases terrigènes
80 : DL	Détritique du large
250 : Côte	Terres émergées

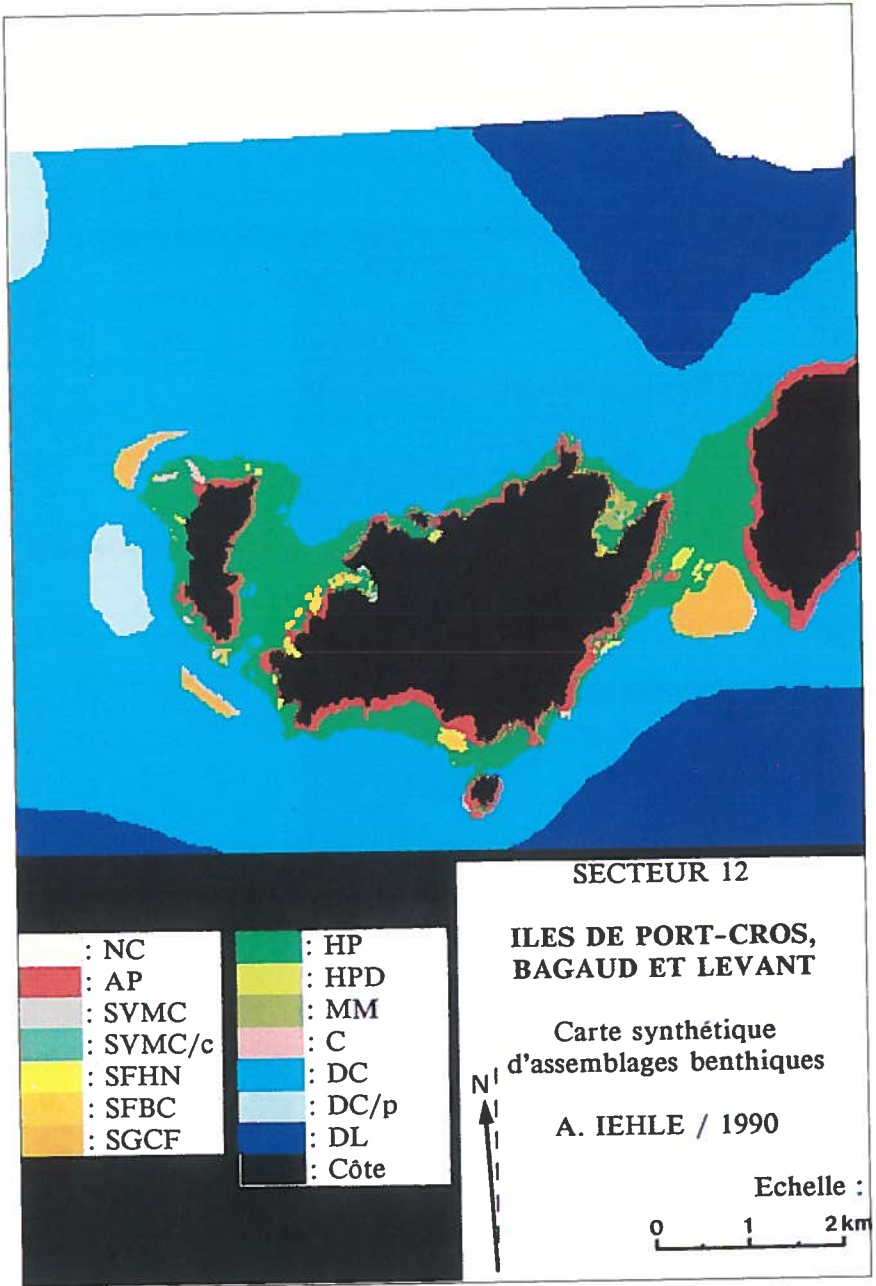


Fig. 2 : Carte synthétique du secteur n° 12 (îles de Port-Cros, Bagaud et du Levant).

Une fois la base de données réalisée, on réalise pour chaque secteur une carte synthétique dont la conformité à la réalité sera meilleure que pour chacune des cartes utilisées pour son élaboration. La méthode utilisée fait appel au formalisme des ensembles flous et est décrite en détail dans MEAILLE (1988). A chaque carte est associée une image représentant le degré de confiance (fiabilité) dans l'information de chaque pixel de cette carte. La carte synthétique (Fig. 3) et construite

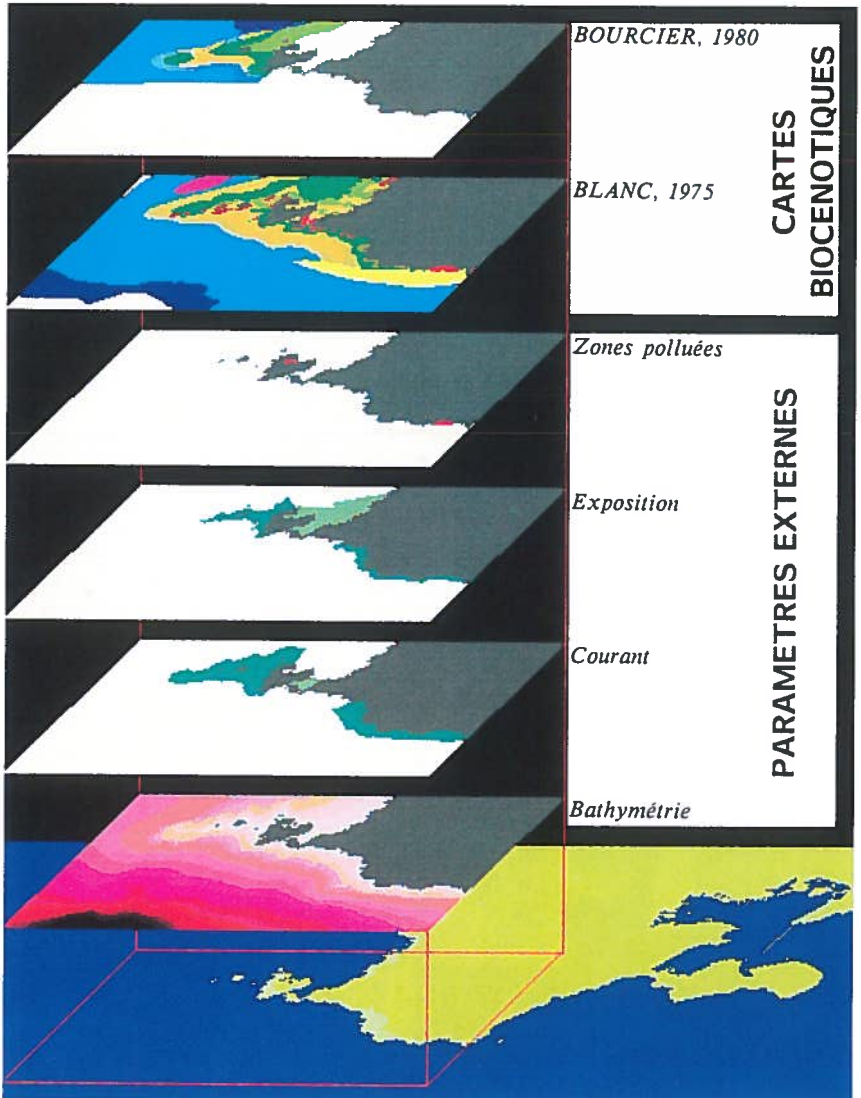


Fig. 3 : Structure multicouche de la base de données.
Secteur n° 4 (Cap Sicié et île des Embiez. Même légende que Fig. 2.

pixel par pixel en utilisant l'information qui a la plus grande fiabilité. En cas de fiabilités de valeurs égales, la préséance est accordée à la carte la plus récente. Cette carte de synthèse est accompagnée d'une carte dite de désaccord représentant en chaque pixel la nature du conflit entre la classe utilisée pour la synthèse et la classe la plus fréquente pour ce pixel donné. Cette carte des désaccords permet la critique de la carte de synthèse en soulignant les zones de conflit entre les auteurs des cartes originales.

Tableau II— Cartes utilisées pour le secteur n° 12 (Port-Cros).

REFERENCE	ZONE CARTOGRAPHIEE
PERES et PICARD 1963	Port-Cros et Levant
TAILLEZ 1966	Port-Cros et Bagaud
AUGIER et BOUDOURESQUE 1967	Baie de la Palud
AUGIER et BOUDOURESQUE 1970a	Baie de Port-Man
AUGIER et BOUDOURESQUE 1970b	Baie de Port-Cros
BLANC 1975	De la presqu'île de Giens au cap Camarat
AUGIER et BOUDOURESQUE 1976	Anse des curés Pointe de la Palud Pointe de la Marma Pointe de la Galère Baie de Port-Man Pointe de Port-Man Cap du Tuf Pointe du vaisseau Pointe du vallon
BLANC et JEUDY de GRISSAC 1978	Passe des grottes Passe de Port-Cros

Les cartes utilisées dans la synthèse du secteur n° 12 (Port-Cros) sont notées dans le tableau II. Par souci de lisibilité après reproduction du document original, les classes AP et AP/e ont été réunies en «Algues Photophiles» (AP).

Les deux bandes de SGCF situées au Sud et au Nord - Nord-Est de l'île de Bagaud représentent l'artéfact décrit par MEAILLE *et al.* (1988). Ces parcelles de SGCF, peu fiables, font partie d'une carte de fiabilité médiocre qui n'est qu'en partie recouverte par une ou plusieurs cartes plus fiables mais moins étendues géographiquement.

Outre les cartes de peuplements benthiques, la base de données contient des cartes dites de paramètres (Fig. 2). Quatre paramètres ont été retenus parmi ceux désignés par les spécialistes consultés comme étant potentiellement déterminants pour la répartition des différentes classes répertoriées. Il s'agit de la profondeur (bathymétrie), de la localisation des baies et caps sous-marins (exposition), de la localisation des courants, et enfin de celle des ports et zones polluées (pollution).

Les cartes bathymétriques sont issues des cartes IGN au 1/25 000ème qui ont été numérisées. Une interpolation linéaire entre les isobathes a été réalisée à la société SODETEG-T.A.I. afin de disposer d'une bathymétrie avec un pas de un mètre.

Les cartes de paramètres présentent une différence essentielle avec les cartes des peuplements benthiques: on suppose que les informations qu'elles contiennent sont vraies. En effet, ces cartes qui ne sont pas basées sur une visualisation du benthos ne représentent, pour l'instant, que les conditions physiques du milieu étudié. Il est à noter que, si les informations ainsi stockées sont annexes au thème principal du SIG, elles peuvent cependant se révéler intéressantes à posséder dans le cas d'une étude dont l'envergure dépasserait le sujet des assemblages benthiques. Les cartes de paramètres constituent donc une banque de données classique dans laquelle chaque canal représente un thème distinct.

Voyons maintenant de quelle façon ces informations sont stockées, et principalement comment définir le degré de résolution de la base de données. Si la résolution était parfaite, on retrouverait en tout point de la carte numérisée la même information que celle présentée par l'homologue de ce point sur la carte originelle. Ce point théorique n'ayant pas de réalité géographique, en pratique la résolution de la base de données sera inférieure à celle de la carte, entraînant une perte d'information par sous-échantillonnage. Ainsi sur la figure 4, l'information sur la carte initiale «B présente l'information 2» est perdue avec les deux tailles de pixels représentées.

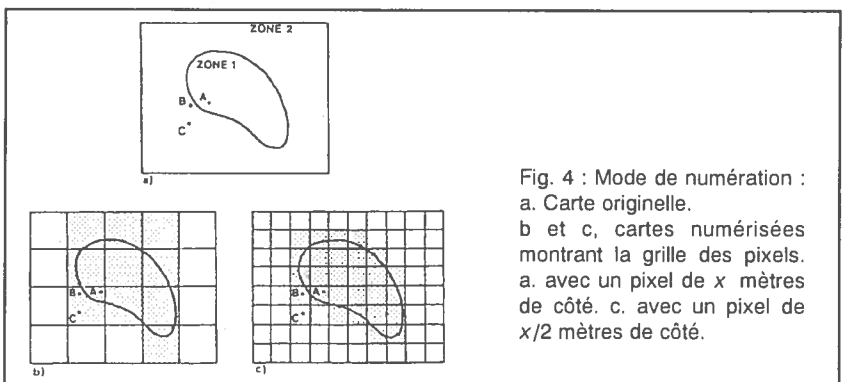


Fig. 4 : Mode de numérisation :
a. Carte originelle.
b et c, cartes numérisées montrant la grille des pixels.
a. avec un pixel de x mètres de côté. c. avec un pixel de $x/2$ mètres de côté.

Par ailleurs, l'ordre dans lequel seront numérisées les différentes classes de peuplements benthiques intervient comme le montre la figure 5. Sur cette figure, si la zone 1 est numérisée en premier, les points A et B sur la carte numérisée auront la même valeur 1 d'information. Dans le cas contraire, ils présenteront l'information 2.

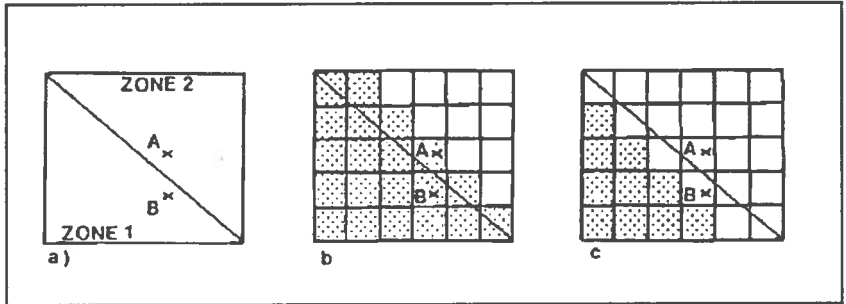


Fig. 5 : Préséance à la numérisation : a. Carte originelle. b. Zone 1 (grisés) numérisée en premier. c. Zone 2 (blanc) numérisée en premier.

Le choix de la taille du pixel est donc une étape fondamentale de la constitution d'une base de données. L'utilisation d'un pixel de 25 mètres x 25 mètres pour cette étude est le résultat d'un compromis entre les deux paramètres: Objectifs du SIG et Information disponible. L'objectif principal de l'implémentation de ce système a été, en plus des intérêts de l'archivage et du suivi temporel, de réaliser une cartographie la plus fiable possible (et de toutes façons plus fiable que les cartes dont elle est issue) des peuplements benthiques sur une zone dont les limites ne sont pas figées. La surface actuellement couverte est à l'échelle d'un département, et de ce fait, cette banque de données n'est en aucun cas destinée à être utilisée pour des travaux de faible étendue spatiale (comme une plongée par exemple).

Il est évident que les impératifs de résolution au sol dépendent du type d'objet observé. Dans notre cas, une biocénose comme celle du Détritique Côtier (DC), présentant une faible variabilité spatiale et une étendue généralement importante, sera correctement décrite en utilisant un pixel de 25 mètres voire de 50 mètres de côté. A l'opposé, un pixel de 25 mètres ne pourra rendre compte des variations spatiales de certains assemblages, comme ceux de la roche infralittorale, qui s'observent sur des distances de l'ordre du décimètre. C'est pour cette raison que seuls trois faciès ont été utilisés pour le groupe des Algues Photophiles alors que MEINESZ *et al.* (1983) en répertorient treize.

Le deuxième point à intégrer est la précision (qu'il ne faut pas confondre avec la fiabilité) de l'information dont on disposera pour constituer la base de données. Les documents existants sont des croquis et des cartes dont les échelles varient du 1/1 000ème au 1/15 000ème. Si les premiers présentent une grande finesse de la maille

d'observation, ils ont par contre souvent peu de valeur sur le plan du repérage géographique. Les seconds, facilement repérables dans l'espace, n'ont de valeur informationnelle que pour les classes couvrant de grandes surfaces.

RESULTATS ET DISCUSSION

Nous avons vu que le milieu marin auquel s'applique notre SIG nécessite la prise en compte de l'inexactitude potentielle de l'information car la redondance qui existe entre les différents canaux d'une même image nécessite un choix pour la constitution d'une seule carte synthétique.

Examinons les sources d'erreurs dans notre système, en distinguant, à la suite de WALSH *et al.* (1987), une «erreur intrinsèque» et une «erreur opérationnelle». La première est l'erreur contenue dans les données et la seconde celle liée à l'utilisation de ces données dans le système.

FACTEURS INFLUANT SUR L'ERREUR INTRINSEQUE

Age de la carte

La probabilité d'une évolution de la répartition des peuplements étudiés croît avec l'âge de la carte, augmentant ainsi le risque d'erreur d'information en un pixel donné. Il est à noter que le vieillissement de la carte n'a pas une égale influence sur tous les codes utilisés. En effet, un herbier de posidonies peut rapidement évoluer en une matie morte de posidonies ou en sables grossiers sous l'influence d'un courant de fond alors que le passage d'une zone d'algues photophiles en un autre code est beaucoup moins probable.

Identification de l'information

La reconnaissance erronée d'une classe peut se produire lorsque les conditions d'utilisation de la méthode de cartographie employée ne sont pas respectées (par exemple : trait de drague effectué à vitesse trop élevée, photo aérienne prise avec une mauvaise visibilité ou dans la zone de réflexion spéculaire de la lumière solaire), ou que les limites d'application de cette méthode sont dépassées (par exemple : cartographie des fonds à une profondeur supérieure à 35 mètres par photo-interprétation). Ce type d'erreur dépend également de l'expérience de l'auteur de la carte (par exemple : prélèvement dans

lequel les espèces caractéristiques d'une biocénose ne sont pas toutes présentes, ou sont en mélange avec celles d'autres biocénoses).

Localisation de l'information récoltée

L'erreur sur la localisation géographique de l'information peut se produire en photo-interprétation lorsque les documents utilisés n'ont pas subi de correction de la parallaxe, ou pour un draguage au cours duquel l'engin a échantillonné le sédiment de manière inconstante sur son trajet ou encore lorsque le positionnement de l'observateur (embarqué ou en plongée) est inexact.

Dans un SIG, la précision de la position d'une observation est tout aussi importante que la nature de l'observation elle-même. En effet, si la localisation de l'information lors de sa collecte est inexacte, l'étape de numérisation attribuera cette observation à un pixel ne correspondant pas à la position réelle de l'observation. Dans le meilleur des cas, cette erreur n'aura aucune répercussion sur l'exactitude de la carte si les deux pixels en cause devaient avoir le même code et on peut penser qu'au pire, l'information sera quand même présente, bien que légèrement décalée. Mais cette opinion globaliste qui considère l'information totale contenue dans la carte ne correspond en rien à la méthode d'identification de l'information par le système. En effet, dans le cas de la fonction fondamentale du système (la création d'une carte synthétique), l'information globale n'existe pas. Le pixel, unité élémentaire de surface est aussi une unité élémentaire d'information totalement indépendante du reste de la carte. On retrouve ici la structure réelle de l'image : une matrice de ixj valeurs qui ne seront comparées qu'avec celles occupant la même position dans les autres canaux utilisés.

Tracé de la carte

Cette erreur peut se scinder en :

- une erreur de positionnement de l'information observée dont les conséquences seront plus ou moins importantes selon l'échelle de la carte. Celle-ci s'applique à toutes les techniques de cartographie ;
- une erreur due à l'extrapolation des informations, qui ne concerne pas les cartes dressées par photo-interprétation. Les cartes dont les données sont des observations ponctuelles du terrain, sont créées par extension (en fonction des connaissances de l'auteur sur la zone cartographiée) de l'information de ces points aux points voisins. Bien entendu, l'erreur de cette extension d'information sera d'autant plus réduite que la maille des observations sera serrée.

ERREURS OPERATIONNELLES

Erreurs dues à la résolution de numérisation

Nous avons vu précédemment que la numérisation d'une carte peut créer des imperfections aux contours des zones représentées, en supposant que ces contours sont exacts sur la carte, ce qui n'est pas toujours le cas, selon que ceux-ci proviennent d'une extrapolation ou d'une photo-interprétation.

Erreur de codage

L'étape de codage de l'information récoltée est une étape délicate et évidemment décisive (CAMPBELL et MORTENSON, 1989). En fonction de l'échelle de la carte, les peuplements décrits sont des regroupements ou des scissions des classes définies pour notre banque de données. De plus, pour les cartes les plus anciennes, chaque auteur a un vocabulaire qui lui est propre et qu'il faut normaliser. Lorsqu'ils sont rencontrés, les faciès qui n'ont pas été distingués dans la table des codes utilisés sont rattachés à la biocénose typique dont ils sont issus. La perte d'information qui en découle est acceptable car toutes les données non répertoriées par la table des classes ne sont pas utilisables à l'échelle choisie pour cette étude. Par contre, lorsque la classification employée par l'auteur de la carte est moins précise que celle de notre étude, les zones délimitées regroupent alors, plusieurs classes et le codage est beaucoup moins aisé. Ainsi, si les classes associées dans une même zone appartiennent à des groupes distincts, cette zone est considérée comme non cartographiée. De la même façon, dans les zones de transitions entre deux biocénoses bien déterminées, il est souvent difficile de choisir entre les deux codes des classes en présence.

Erreur à la numérisation

Cette erreur peut se différencier en une erreur inhérente à la numérisation et une erreur de manipulation lors de la numérisation.

La première vient du fait que les zones à délimiter ne sont pas des polygones dont les côtés se couperaient à angle droit ; ainsi les limites de ces zones ne correspondent-elles pas à celles de la grille des pixels de l'image. La figure 5 b permet de mieux saisir le problème. La numérisation de la zone 1 en premier entraîne une erreur majorante sur la surface de la parcelle 1 au détriment de la parcelle 2.

L'erreur de manipulation correspond à l'écart accidentel que l'on peut faire lors du suivi de la limite de zone numérisée avec la souris. Cette déviation est inférieure à 1 mm sur la carte numérisée pour un opérateur attentif, mais l'erreur résultante sur la carte stockée dépend de l'échelle du document original utilisé. Dans notre cas, pour lequel un

pixel a 25 mètres de côté, si l'échelle originale à la numérisation est de 1/25 000ème, un écart de 1 mm correspondra à une erreur de 1 pixel sur la carte numérique. Cette erreur sera quatre fois plus importante pour une échelle de numérisation de 1/100 000ème. Toutefois, cette erreur pourra être corrigée lors de l'analyse sur écran de la carte numérisée.

Approximations lors de la rectification géométrique

Lorsque la carte originelle à numériser a une grande échelle, le nombre de points remarquables (amers) présents à la fois sur cette carte et sur celle du trait de côte-référence IGN est faible. Le modèle polynomial de rectification est donc calculé sur un nombre restreint de points, ce qui augmente la probabilité d'un décalage qui peut aller jusqu'à trois pixels entre la côte-référence et celle de la carte rectifiée. Le problème est le même pour les cartes dont la morphologie de la côte est monotone. Il faut bien entendu ajouter à ces deux cas celui des cartes dont le trait de côte, indépendamment du nombre d'amers utilisables, ne ressemble que très vaguement à la référence.

SOLUTIONS AUX ERREURS INTRINSEQUES

Les cartes de Fiabilité

Les erreurs intrinsèques que nous avons détaillées ne peuvent bien entendu pas être réduites ou annulées. Il est de plus impossible en pratique de les évaluer individuellement. La méthode retenue consiste donc à attribuer en chaque point de la carte une estimation de la confiance dans son information. Cette mesure de fiabilité doit être comprise au sens de mesure de la conformité à la réalité. Elle correspond donc à la certitude (confiance maximale) diminuée d'une estimation de la valeur globale des six types d'erreurs intrinsèques. Les cartes de fiabilité prises individuellement sont un outil intéressant pour la critique de la qualité d'une carte, mais c'est lors de la comparaison des différentes cartes de la même zone que ces documents seront essentiels. En effet les différentes cartes existantes apportent des informations redondantes et bien souvent conflictuelles. Les cartes de fiabilité vont nous permettre de gérer ces désaccords d'informations.

Création manuelle des cartes de fiabilité

Une valeur de fiabilité prise sur une échelle arbitraire (Tableau III) est attribuée en chaque point de la carte originale par un expert en écologie marine méditerranéenne. Ce spécialiste tient compte pour cela des paramètres (i) âge de la carte, (ii) méthode de cartographie, (iii) nom

de l'auteur, (iv) échelle de la carte, ainsi que (v) de ses connaissances personnelles de la zone décrite. Si cette méthode est simple et rapide, son formalisme reste imprécis et la comparaison des cartes de fiabilité ainsi produites est donc discutable.

Tableau III — Quelques coefficients de fiabilité et leurs significations. Des valeurs intermédiaires sont utilisées pour les cas intermédiaires. D'après MEAILLE *et al.* (1988).

COEFFICIENT DE FIABILITE (CN _f)	SIGNIFICATION
0	Non cartographié
4	Fiabilité nulle : limite placée de façon totalement empirique et sans aucune justification.
8	Fiabilité approximative : limite placée de façon empirique, mais en fonction d'une certaine logique. Ex : dans un secteur donné, la limite inférieure de l'herbier se situe en général entre 30 et 40m : on a donc suivi les isobathes.
12	Fiabilité moyenne : limite placée en fonction de données concrètes ; l'imprécision vient du type de données ou du repérage.
16	Fiabilité bonne : limite placée en fonction de données concrètes; maillage serré (1 observation/100m) ; données très précisément localisées (ex : sonar latéral, transects balisés, photo-interprétation avec correction de la parallaxe).
20	Totalement fiable : l'erreur est inférieure ou égale au pixel.

Classification au moyen des tables de hash

Nous avons donc étudié une méthode objective d'estimation de la fiabilité dérivée de la télédétection. Soit n le nombre de paramètres qui seront utilisés pour la classification. A chaque pixel, on peut associer le n -uplet des valeurs prises dans les n canaux de paramètres. Par analogie avec la télédétection, cette combinaison de valeurs des variables externes sera appelée la signature du pixel. A l'ensemble des pixels que constitue une zone géographique, correspond donc un nuage de signatures plongé dans l'espace à n dimensions des paramètres. Les tables de hash (de l'anglais «hash» : hacher) constituent un mode de stockage (et de gestion) de cette information multidimensionnelle particulièrement performant. Une table de hash (encore appelée table

d'association biunivoque) est une structure contenant la liste exhaustive des signatures rencontrées dans une partie de l'image. La partie optimisation de l'adressage est transparente à l'utilisateur qui se contente donc d'utiliser les tables de hash comme des histogrammes croisés à n entrées (BOSSUT, 1986). En chaque enregistrement de la table sont déterminées les valeurs suivantes : les coordonnées multispectrales du point, le nombre d'occurrences, c'est-à-dire le nombre de pixels de la zone étudiée présentant cette signature, la classe, c'est-à-dire le code qui sera restitué à tous les pixels présentant la même signature spectrale. Ce code est calculé de façons diverses selon ce que l'on veut lui faire décrire. Dans cette étude il correspond à la fiabilité de la signature.

Les paramètres utilisés dans cette étude ont été introduits précédemment (bathymétrie, localisation des ports, localisation des zones caractérisées par un mode exposé ou un mode calme, et localisation des zones soumises à un courant important). Pour chaque classe x d'une carte de biocénoses, on définit la signature de chaque pixel appartenant à cette zone. On construit donc la table de hash de cette carte pour la classe considérée, qui décrit la dispersion de la classe x dans l'espace des paramètres. Cette technique s'apparente à la constitution de polygones d'apprentissage pour une classification supervisée. En pratique, on utilisera les signatures de la classe x, non pas sur une seule carte, mais sur l'ensemble des cartes afin d'avoir la plus grande information possible (i.e. la dispersion maximale de la classe dans l'espace des paramètres). Il en résulte une table constituée par toutes les signatures rencontrées sur les pixels de la classe x de toutes les cartes utilisées.

Le but est de déterminer une mesure de la fiabilité de l'appartenance de chaque pixel à la carte à laquelle il est associé. Nous allons pour cela utiliser la seule information disponible qui est celle du nombre d'occurrences de chaque signature. A partir des occurrences, on détermine la plausibilité de chaque combinaison de valeur des quatre paramètres. Pour cela, les taxons utilisés sont alors considérés comme des ensembles flous (ZADEH, 1965 ; PRADE, 1982). En logique Booléenne (classique), si un point appartient à un ensemble A, il ne peut appartenir à son complémentaire dans l'espace considéré. En logique floue, il peut exister, entre un ensemble et son complémentaire, une zone de chevauchement due à l'imprécision de leurs limites.

D'après PRADE (1982), un ensemble flou A est défini par un référentiel X, et une fonction caractéristique ou fonction d'appartenance μ_A :

$$\mu_A : X \in X \rightarrow \mu_A(X) \in [0,1]$$

Dans notre cas, μ_A est appelé fiabilité ou plausibilité et sera exprimée en fonction du nombre d'occurrences :

$$\mu_A(X_i) = ip(X_i) + \sum_{j=1+i}^n p(X_j) \quad i = 1, \dots, n$$

où $p(x_i)$ est la probabilité d'appartenance de x_i à A (i.e. la probabilité de la signature de rang i), les x_i étant rangés dans le sens des valeurs décroissantes de probabilité.

Chaque signature spectrale va alors recevoir un code proportionnel à sa plausibilité. Ce codage s'effectue sur des valeurs allant de 0 à 100 permettant ainsi d'obtenir la valeur de fiabilité sous forme de pourcentage.

Les plausibilités respectent le même ordre que les probabilités. Quel est donc leur avantage sur ces dernières ?

On constate (PIVAUDRAN, comm. pers.) que pour les taxons hétérogènes vis-à-vis de leurs signatures spectrales (c'est notre cas), les distributions de probabilité ont tendance à présenter une allure uniforme. Et si ce n'est pas une règle générale, on ne peut se permettre de prendre l'hypothèse d'une distribution Gaussienne. Si l'on codait les signatures proportionnellement aux probabilités, on obtiendrait des valeurs plus ou moins élevées selon le nombre de signatures du taxon, et à l'intérieur d'un taxon, ces codes (que nous utilisons comme mesure de la fiabilité) seraient tous plus ou moins équivalents du fait de la distribution uniforme. La fiabilité serait donc plus dépendante du nombre de signatures contenues dans le taxon que de la fréquence de ces signatures. L'intérêt de l'utilisation des plausibilités réside dans leur caractère normatif. En effet, le point de plus forte probabilité reçoit la plausibilité 1 et les points suivants ont une plausibilité dont la différence avec celle du point précédent est proportionnelle à leur écart en probabilité. L'effet «taille de l'échantillon» (qui influe sur le nombre de signatures rencontrées) s'en trouve donc relativisé. Les codes de fiabilités seront donc comparables, au sens de leur représentativité de la fréquence des signatures des taxons. Une fois ce codage effectué, on restitue le résultat en attribuant à chaque pixel de la classe x (et pour toutes les cartes) le code correspondant à sa signature spectrale. On obtient ainsi, pour les pixels de la zone recodée, une carte de plausibilité d'appartenance à la classe x .

Si l'on étend ce recodage à l'ensemble de l'image, on obtiendra en tout pixel sa fiabilité d'appartenance à la classe x , et ceci même si l'auteur de la carte lui avait attribué une autre classe. On pourra ainsi déterminer en un pixel sa fiabilité d'appartenance à la classe sous laquelle il est annoncé mais également à chacune des autres classes

répertoriées. Ce concept apparaît très intéressant car il rejoint la notion de zone de transition pour laquelle un pixel appartient aux deux biocénoses impliquées. La figure 6 donne le résultat du hashage sur le secteur n° 4 (îles des Embiez). On observe sur la carte un aspect strié des valeurs de fiabilité. Ce phénomène est dû à la nature des quatre paramètres choisis. En effet, les localisations des zones polluées, des zones de courant et des modes d'exposition sont codées sur deux ou trois niveaux (exemple : pas d'information, mode battu, mode calme) alors que nous avons conservé pour la variable profondeur les 202 niveaux issus de l'interpolation des isobathes. Il y a donc eu constitution d'une signature, et donc calcul d'un code de fiabilité, chaque fois que l'analyse rencontrait une nouvelle valeur de profondeur.

On observe (Fig. 6c) que globalement, la carte produite par l'utilisation des tables de hash est assez semblable à celle issue de la méthode manuelle. Cependant, l'attribution des fiabilités par les tables de hash n'élimine pas le problème souligné par MEAILLE (1988) : le recouvrement par une classe de bonne fiabilité d'une classe de moindre fiabilité mais représentant une surface plus importante. Les débordements de cette dernière classe qui en résultent constituent des schémas de répartition très improbables. Ainsi, sur la figure 6 c, la synthèse obtenue par la classification de l'image présente, au-delà des zones de DC/m et DC/c, des bandes de SFBC qui sont sans aucun doute des artéfacts. Néanmoins, ce type d'imperfection tendra à disparaître avec l'augmentation de la quantité d'informations stockées. La technique de classification que nous avons présentée constituera alors une vraie base de connaissances sur la répartition des assemblages benthiques méditerranéens. La supériorité de cette méthode sur la méthode manuelle sera alors l'objectivité de son mécanisme mathématique.

SOLUTIONS DES ERREURS OPERATIONNELLES

L'erreur inhérente à la numérisation est difficile à réduire. Ceci supposerait des méthodes de préséance à la numérisation aussi complexes que fastidieuses à réaliser.

L'erreur de manipulation est diminuée en utilisant, lors de la numérisation, les documents-sources à une échelle d'au moins 1/1 250ème pour laquelle la déviation maximale de numérisation de 1 mm correspond à 1/2 pixel. Nous avons considéré cette déviation maximale comme acceptable.

L'erreur due à la rectification géométrique, qui n'est pas modifiable, pourrait être quantifiée en attribuant à la carte une valeur de «fiabilité de géométrie» qui dépendrait de la qualité (au sens de la conformité de la carte au trait de côte référence) du modèle de rectification géométrique utilisé.

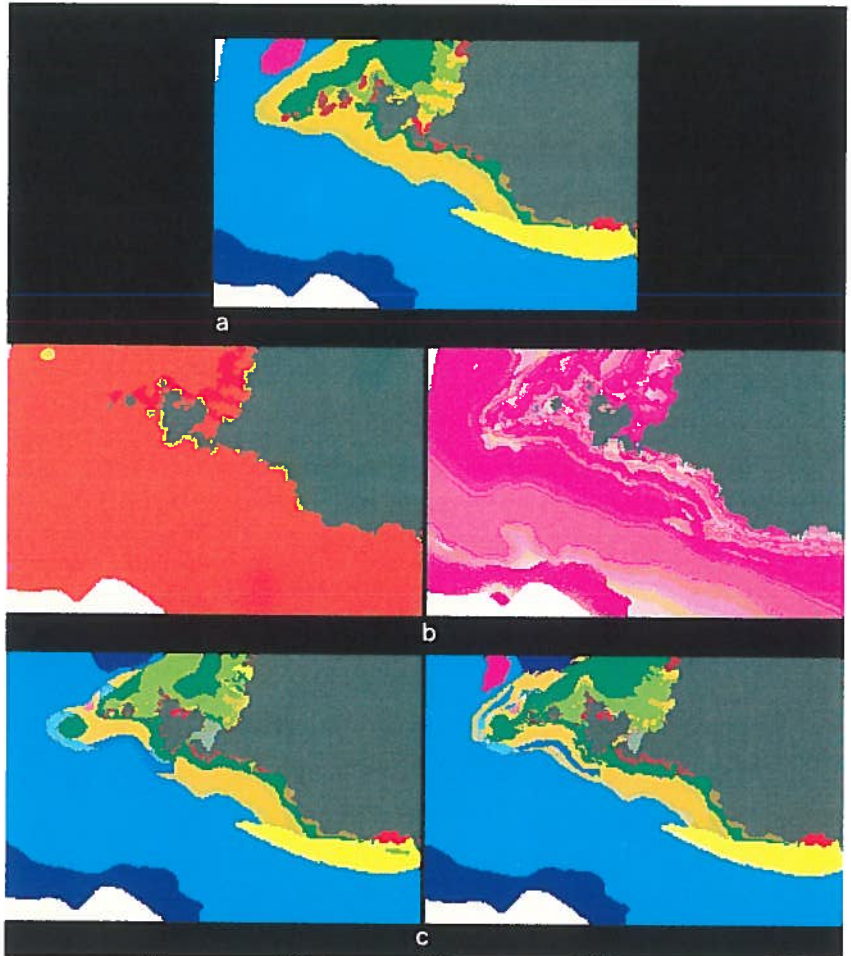


Fig. 6 : Comparaison des méthodes d'attribution des fiabilités, manuellement et par classification d'image (Secteur 4 : île des Embiez).

a. Carte de BLANC (1975) originelle. Légende Fig. 2.

b. Cartes des fiabilités à gauche par la méthode manuelle, à droite par les tables de hash, sur la carte de BLANC (1975).

La table de couleurs est linéaire. La fiabilité est meilleure pour les teintes sombres.

c. Cartes de synthèse de tous les documents disponibles pour le secteur n° 4, à gauche par la méthode manuelle, à droite par les tables de hash.

CONCLUSION

La nécessité grandissante de gérer les fonds littoraux, pour leur protection comme leur exploitation, comme on le fait pour les surfaces émergées, souligne la difficulté d'appréhender ce milieu masqué par la couche d'eau qui le recouvre.

Les SIG apparaissent comme les seuls moyens d'obtenir une réponse précise à des questions de plus en plus complexes se rapportant à la géographie. Le SIG présenté dans ce travail fait partie de ces systèmes d'aide à la recherche et à la décision. En effet, si les questions dont nous avons débattu s'appliquent à la restitution d'une carte synthétique fiable, celle-ci doit être considérée comme un outil de base pour des études très diverses. Tel qu'il est actuellement, le SIG des peuplements benthiques «MBA» est un outil très efficace pour les problèmes de filtrage :

- spatial (déterminations multicritères de zones à protéger, mise en évidence de zones propices à un aménagement du littoral, évaluation des surfaces occupées selon divers paramètres ou concernées par de possibles modifications),
- ou temporel (visualisation de l'évolution au cours du temps des surfaces respectives des différentes classes répertoriées).

Le problème de la détermination du degré de confiance à chacune des étapes du système est crucial pour la critique et l'utilisation raisonnée du document final. La technique de classification pour l'attribution des fiabilités que nous avons introduite présente quelques faiblesses liées à la nécessité d'une grande quantité d'information pour obtenir la dispersion maximale dans l'espace des paramètres. De plus, elle n'admet pas les facteurs non géographiques (hormis l'âge de la carte qui est pris en compte lors de la synthèse cartographique) pourtant influant sur la fiabilité, tels que la méthode de collecte des données ou l'échelle originelle. Cependant, cette méthode apporte l'automatisation de l'étape complexe de l'attribution des fiabilités et permet d'obtenir des résultats reproductibles et comparables.

REMERCIEMENTS

Les auteurs tiennent à remercier O. PIVAUDRAN pour l'adaptation de l'outil informatique au problème étudié, ainsi que S. RUECH et I. GEIGER pour l'interpolation des données bathymétriques.

BIBLIOGRAPHIE

- AUGIER H., BOUDOURESQUE C.-F., 1967. — Végétation marine de l'île de Port-Cros (Parc National). I : la baie de la palud. *Bull. Mus. Hist. nat. Marseille, Fr.*, 27 : 93-124, 1 pl. h.t.
- AUGIER H., BOUDOURESQUE C.-F., 1970 a. — Végétation marine de l'île de Port-Cros (Parc National). V : la baie de Port-Man et le problème de la régression de l'herbier de Posidonies. *Bull. Mus. Hist. nat. Marseille, Fr.*, 30 : 145-164.
- AUGIER H., BOUDOURESQUE C.-F., 1970 b. — Végétation marine de l'île de Port-Cros (Parc National). VI : le récif-barrière de posidonies. *Bull. Mus. Hist. nat. Marseille, Fr.*, 30 : 221-228.
- AUGIER H., BOUDOURESQUE C.-F., 1976. — Végétation marine de l'île de Port-Cros (Parc National). XIII : documents pour la carte des peuplements benthiques. *Trav. sci. Parc nation. Port-Cros, Fr.*, 2 : 9-21, 1 pl. h.t.
- BLANC J., 1975. - Recherches de sédimentologie appliquées au littoral rocheux de la Provence. Aménagement et protection. *Centre National Exploitation Océans, Fr.* : 164 T.
- BLANC J.J., JEUDY DE GRISSAC A., 1978. — Recherches de géologie sédimentaire sur les herbiers à Posidonies du littoral de la Provence. *Centre National Exploitation Océans, Fr.* : i-vii, 1-185, 42 pl. h.t.
- BOSSUT P., 1986. - Analyse des données : — Application à l'analyse automatique d'images multispectrales. *Thèse Docteur-Ingénieur, Ecole Nationale Supérieure des Mines de Paris, Paris, Fr.* : 290 p.
- BOUDOURESQUE C.-F., MEINESZ A., LEFEVRE J.-R., 1985. — Cartographie des peuplements benthiques marins de la Corse : I. La formation récifale à *Posidonia oceanica* de Saint-Florent. *Ann. Inst. oceanogr.*, Fr., 61(1) : 27-38.
- CAMPBELL W.G., MORTENSON D.C., 1989. — Ensuring the quality of geographic information system data : A practical application of quality control. *Photogrammetric Engineering Remote Sensing*, 55 (11) : 1613-1618.
- COLANTONI P., GALLIGNANI P., FRESI E., CINELLI F., 1982. — Patterns of *Posidonia oceanica* (L.) Delile beds around the island of Ischia (gulf of Naples) and adjacent waters. *Mar. Ecol.*, Germ., 3(1) : 53-74, 2 pl. h.t.
- JEUDY DE GRISSAC A., MEINESZ A., BOUDOURESQUE C.-F., ASTIER J.-M., BOURCIER M., LEFEVRE J.-R., 1985. - Localisation de l'herbier de Posidonies sur le littoral P.A.C.A. Etat des connaissances. *GIS Posidonies et DRAE PACA, Marseille, Fr.*, 1-22 + 1-37 + 1-19.
- MEAILLE R., 1988. — Les systèmes d'information géographique : Structure, mise en œuvre et utilisation dans différentes études. *Thèse Doctorat, Université de Nice, Fr.* : 182 p.
- MEAILLE R., WALD L., BOUDOURESQUE C.-F., 1988. — Cartes des peuplements benthiques en Méditerranée: constitution d'une banque de données. *Oceanol. Acta, Fr.*, 11(2) : 201-211.

- MEINESZ A., BOUDOURESQUE C.-F., FALCONETTI C., ASTIER J.M., BAY D., BLANC J.J., BOURCIER M., CINELLI F., CIRIK S., CRISTIANI G., DI GERONIMO I., GIACCONE G., HARMELIN J.G., LAUBIER L., LOVRIC A.Z., MOLINIER R., SOYER J., VAMVAKAS C., 1983. - Normalisation des symboles pour la représentation et la cartographie des biocénoses benthiques littorales de Méditerranée. *Ann. Inst. océanogr., Paris, Fr.*, 59 (2) : 155-172.
- MEINESZ, A., CUVELIER, M., LAURENT, R., 1981. — Méthodes récentes de cartographie et de surveillance des herbiers de Phanérogames marines. Leurs applications sur les côtes françaises de la Méditerranée. *Vie Milieu, Fr.*, 31(1) : 27-34.
- PERGENT G., PERGENT C., 1988. — Localisation et état de l'herbier de Posidonies sur le littoral PACA : Bouches du Rhône. *DRAE. PACA et GIS Posidonie Edit., Marseille, Fr.*, 1-53.
- PRADE 1-1., 1982. — Modèles mathématiques de l'imprécis et de l'incertain en vue d'applications au raisonnement naturel. *Thèse Doctorat, Université Paul Sabatier Toulouse, Fr.*: 234 P.
- RAMOS-ESPLA, A.A., 1 985. — Contribucion al conocimiento de las biocenosis bentonicas litorales de la Isla Plana o Nueva Tabarca (Alicante). *La reserva marina de la Isla Plana o Nueva Tabarca (Alicante)*, RAMOS Edit., : 111 147, 1 pl. h.t.
- TAILLIEZ P., 1966. — Carte des peuplements benthiques autour de l'île de Port-Cros. Document inédit, bibliothèque du LBMEB, Faculté des Sciences de Luminy, Marseille, France .
- WALSII S.J., LIGIHTFOOT D.R., BUTLER D.R., 1987. — Recognition and assesment of error in geographic information systems. *Photogrammetric Engineering Remote Sensing*, 53 (10) : 1423-1430.
- ZADEH L.A., 1965. — Fuzzy sets. *Information Control*, 8 : 338-353.