



HAL
open science

Téledétection environnementale en Polynésie française Apport de la classification SVM pour la cartographie du couvert végétal en milieu insulaire tropical

Benoît Stoll

► **To cite this version:**

Benoît Stoll. Téledétection environnementale en Polynésie française Apport de la classification SVM pour la cartographie du couvert végétal en milieu insulaire tropical. Traitement du signal et de l'image [eess.SP]. Université de la Polynésie française, 2014. tel-03777653

HAL Id: tel-03777653

<https://hal.science/tel-03777653v1>

Submitted on 15 Sep 2022

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.

UNIVERSITÉ DE LA POLYNÉSIE FRANÇAISE



HABILITATION À DIRIGER DES RECHERCHES

Téledétection environnementale
en Polynésie française

Apport de la classification SVM pour la cartographie du
couvert végétal en milieu insulaire tropical

Par Dr. Benoît STOLL

Soutenue sur le campus d'Outumaoro, Punaauia – TAHITI

Le 11 juin 2014

Présenté devant le jury composé de :

<i>Jean-Paul Rudant,</i>	<i>Pr.</i>	<i>Université de Marnes-la-Vallée,</i>	<i>Rapporteur</i>
<i>Jocelyn Chanussot,</i>	<i>Pr.</i>	<i>Université de Grenoble,</i>	<i>Rapporteur</i>
<i>Michel Petit,</i>	<i>DR1.</i>	<i>Directeur IRD-France Sud,</i>	<i>Rapporteur</i>
<i>Alban Gabillon,</i>	<i>Pr.</i>	<i>Université de la Polynésie française,</i>	<i>Examineur</i>
<i>Jean-Pierre Barriot,</i>	<i>Pr.</i>	<i>Université de la Polynésie française,</i>	<i>Examineur</i>

« Le sage peut découvrir le monde sans franchir sa porte. »

Lao-Tseu

REMERCIEMENTS

Je voudrais saisir l'occasion qui m'est donnée ici de remercier tous ceux qui ont contribué de près ou de loin à cette aventure d'une douzaine d'années de recherche en Polynésie française.

Tout d'abord, je voudrais remercier mes collègues au sein du laboratoire GePaSUD pour leur soutien implicite ou explicite : avec en tout premier lieu mon collègue Patrick Capolsini pour m'avoir fait confiance lors de mon recrutement en 2001 et pour m'avoir donné les moyens de trouver ma voie ; merci aussi à Sébastien Chabrier pour sa compétence et son dévouement.

Merci à Jean-Pierre Barriot pour avoir accepté la direction de cette HDR mais aussi pour son soutien appuyé.

Je voudrais remercier tout particulièrement tous les étudiants que j'ai pu encadrer en stage ou en thèse, en effet, sans eux ce mémoire n'aurait pu voir le jour. Merci particulièrement à Robin Pouteau pour son travail d'une grande qualité, ses idées et sa grande ouverture d'esprit scientifique. Je souhaite à tous les chercheurs et enseignant-chercheurs d'encadrer un jour un thésard aussi méritant, passionné et travailleur.

Merci aussi à tous les collègues avec qui j'ai pu collaborer, en France ou à l'étranger, pour toutes ces discussions constructives, et pour ces bons moments partagés lors des conférences à l'étranger, contribuant ainsi à les rendre plus conviviales.

Je voudrais remercier tous ceux qui ont contribué financièrement ou matériellement à rendre ma recherche possible : je pense entre autres aux services du territoire et particulièrement au service de l'urbanisme avec son responsable Didier Lequeux, à la délégation à la recherche avec Tea Frogier et Jean-Yves Meyer.

Merci aussi à Frédéric Jacq pour ses discussions passionnantes et l'accès à ses données terrain.

Je ne voudrais pas oublier ma famille, proche ou lointaine qui est le contrepond affectif de tout ce travail : ma compagne Sabine dont le travail de corrections et de relectures a contribué à éclaircir ma pensée et lui donner un lustre que je ne lui soupçonnais pas mais aussi pour son écoute et son soutien indéfectible ; mes deux enfants Mahana et Vetea qui liront peut être un jour ce document ; ma mère Andrée-Mathilde Ethève qui m'a inspiré mon analogie picturale ; mes frères Nicolas-Pierre, Daniel et ma sœur Griotte parce qu'ils sont toujours présents dans mon cœur ; mes tantes Thérèse et Bernadette.

Enfin, merci aux membres du jury de cette habilitation à diriger les recherches et tout particulièrement aux rapporteurs Jean-Paul Rudant, Jocelyn Chanussot et Michel Petit d'avoir bien voulu me faire l'honneur de poser un regard critique sur ce document.

Un remerciement tout particulier à Michel Petit pour son amitié, son soutien, sa confiance et les nombreuses séances d'échanges d'idées qui ont grandement contribué à la finalisation de ce document.

CURRICULUM VITAE

État civil

Nom : Stoll

Prénom : Benoît

Né le 1^{er} mai 1971 à Toulon (Var)

Nationalité : française



Situation professionnelle

Maître de conférences en informatique et télédétection (section CNU 61) à l'université de la Polynésie française depuis août 2001.

Affecté au laboratoire GePaSUD-Géopôle du Pacifique Sud, EA 4238 (anciennement Terre-Océan) de l'université de la Polynésie française depuis août 2001.

Coordonnées professionnelles

Adresse postale : BP 6570 - Faa'a - 98702 - Tahiti – Polynésie française

Mèl: benoit.stoll@upf.pf

Tel : (+689) 803 879 (GMT-10)

Fax : (+689) 803 961

Qualification

17 Mars 2000 : qualifié en **61ème** section CNU.

Titres et diplômes :

Thèse soutenue le 19 Janvier 2000 à l'Université de Toulon et du Var.

Titre : **Optimisation de fonctions de contraste en séparation de sources**

Spécialité : traitement du signal et des images

Jury : P.Y. Arquès, président, G. Alengrin, rapporteur, C. Jutten, rapporteur, J.-C. Pesquet, examinateur, J.-F. Cavassilas, directeur de thèse, E. Moreau, examinateur.

1995-1996 : **DEA Optique Signal et Image** de l'École Nationale Supérieure de Physique de Marseille Option : **traitement d'images**

1994-1995 : **Maîtrise EEA** de l'Université de Toulon et du Var.

1993-1994 : **Licence Ingénierie Électrique** de l'Université de Toulon et du Var.

1989-1991 : **DUT Génie Électrique et Informatique Industrielle** de l'Université de Toulon et du Var option : **électronique**.

1988-1989 : **Baccalauréat série C** (mathématique et physique).

SOMMAIRE

Sommaire	1
Liste des tableaux	3
Liste des figures	5
Liste des acronymes	7
Préambule.....	9
I. Introduction	11
A. Axe de recherche	11
B. Contexte social	12
1. Facteurs contraignants	12
2. Enseignement	12
3. Responsabilités.....	14
4. Recherche	14
a. Encadrements	14
b. Financements.....	17
c. Collaborations	18
C. Contexte géographique.....	18
1. Genèse des îles volcaniques	18
2. Répartition spatiale de la végétation	19
a. Phytogéographie.....	19
b. Effet de foehn	20
c. Structure verticale de la végétation	21
d. Facteurs anthropiques.....	21
e. Espèces indigènes, exogènes et envahissantes	22
3. Typologie de la végétation	22
D. Questionnements et défis scientifiques	24
II. Analyse des données de télédétection satellitaire	29
A. Donnée optique	29
1. Informations spectrales et texturales	29
a. Classification d'images <i>HR</i> de Tahiti.....	29
b. Régression de la forêt de nuages	32
c. Classification d'attributs de textures	33
2. Étude de la cocoteraie des Tuamotu.....	37
a. Dénombrement des cocotiers	38
b. Fouille de données.....	44
B. Donnée radar	47
1. Donnée radar en polarisation linéaire simple	47
a. Classification mono-source	47
b. Un indice de végétation radar.....	51
c. Étude comparative des bandes radar simples	53
2. Donnée radar polarimétrique.....	54
a. Apport des indices polarimétriques.....	54
b. Étude comparative des modes polarimétriques	56
c. Contribution relative des indices polarimétriques.....	59
C. Données ancillaires	61
1. Descripteurs biophysiques.....	61

a.	Facteurs abiotiques	61
b.	Facteurs biotiques.....	61
2.	Application à la modélisation de la niche écologique de <i>Miconia</i>	62
III.	Les algorithmes de classification	65
A.	Préambule.....	65
1.	La classification.....	65
2.	Mesure de la qualité	66
3.	Comparaison multi-échelle	68
B.	L'algorithme <i>SVM</i>	71
1.	Concepts	71
2.	Optimisation des paramètres	72
a.	Par validation croisée	72
b.	Par auto-configuration itérative.....	72
3.	Classification <i>SVM</i> de données radar	76
4.	Vérité terrain	77
IV.	Utilisation conjointe de données hétérogènes	83
A.	Fusion <i>SVM</i> multi-sources	83
1.	Intérêt	83
2.	Fusion <i>SVM</i> sélective	88
B.	Espèces envahissantes et espèces rares	94
1.	Cartographie de <i>Miconia calvescens</i>	94
2.	Niches écologiques d'espèces rares	96
C.	Santé du récif corallien.....	99
V.	Vision synoptique de mon domaine de recherche.....	103
A.	Synthèse	103
1.	Principaux résultats	103
2.	Transversalité	103
3.	Positionnement scientifique et originalité	104
B.	Perspectives	105
1.	Perspectives en télédétection.....	105
a.	Point d'inflexion.....	106
b.	Démocratisation	107
c.	Formation	107
2.	Perspectives sociétales	108
a.	Atouts et menaces.....	108
b.	Secteurs prioritaires.....	108
c.	Enjeux de la géomatique et de la télédétection	109
d.	Aide à la décision	110
e.	Caractéristiques des images	111
3.	Perspectives de recherche.....	111
a.	Capteurs.....	112
b.	Techniques	112
c.	Thématiques	113
d.	SIG versus télédétection.....	114
e.	Études exhaustives versus télédétection.....	114
f.	Sujets de recherche.....	115
	Conclusion.....	117
	Liste des publications	119

LISTE DES TABLEAUX

tableau 1	Analogie entre la peinture et la télédétection	9
tableau 2	Tableau des enseignements	13
tableau 3	Encadrements et publications scientifiques.....	16
tableau 4	Principales îles hautes de Polynésie française.....	19
tableau 5	Typologie de la végétation de Polynésie française	23
tableau 6	Jeu de classes Tahiti	30
tableau 7	Jeu de classes de textures - Moorea	35
tableau 8	Extrapolation finale du nombre de cocotiers sur l'atoll de Tikehau	44
tableau 9	Jeu de classes - Tuamotu	44
tableau 10	Jeu de classes Opunohu - AirSAR	49
tableau 11	Séparabilités de Jeffries-Matusita - Opunohu	50
tableau 12	Matrice de confusion - Opunohu.....	51
tableau 13	Valeur des indices NDVI et SAR _{vi} par classe - Opunohu.....	52
tableau 14	Pertinence de l'ajout d'indices polarimétriques	55
tableau 15	Pertinence de l'ajout de fréquences radar	55
tableau 16	Comparaison avec le cas C _{VV} , L _{VV} , P _{HV}	55
tableau 17	Comparaison globale.....	56
tableau 18	Les 54 indices polarimétriques.....	57
tableau 19	Indices polarimétriques du mode polarimétrie alternée (ASAR).....	57
tableau 20	Indices polarimétriques du mode polarimétrie duale (PALSAR)	58
tableau 21	Indices polarimétriques du mode "π/4"	58
tableau 22	Comparaison des modes polarimétriques complets et partiels (%).....	59
tableau 23	Exemple de matrice de confusion : Pixel [%]	66
tableau 24	Erreurs d'omission et précision « producteur »	67
tableau 25	Erreurs de commission et précision « utilisateur »	67
tableau 26	Caractéristiques des capteurs	69
tableau 27	Temps de calcul pour l'auto-configuration et l'apprentissage SVM.	74
tableau 28	Précisions par classe des classifications de Wishart et SVM.....	76
tableau 29	Max. de vraisemblance vs SVM / pixels purs vs pixels écotones.....	80
tableau 30	Jeu de classes Nuku-Hiva.....	84
tableau 31	Précisions des classifications mono-sources et fusion SVM – Nuku-Hiva	85
tableau 32	Précisions des classifications mono-sources et fusion SVM - Moorea.....	87
tableau 33	Précisions des classifications mono sources - Moorea.....	90

tableau 34	Précisions par classe de la fusion sélective SVM	91
tableau 35	Précisions de la fusion SVM vs fusion SVM sélective.....	91

LISTE DES FIGURES

figure 1	Le taureau de Picasso (1945)	10
figure 2	Évolution des îles coralliennes	18
figure 3	Influence des vents dominants - Tahiti	20
figure 4	Les hauteurs de Tahiti vues depuis la pointe Venus (Sylvain 1960)	21
figure 5	Masques de l'île de Tahiti	30
figure 6	Classification par maximum de vraisemblance de Tahiti (SPOT5 2002-2009)...	31
figure 7	Régression de la végétation orophile de Tahiti entre 2010 et 2150	33
figure 8	Influence du nombre de vignettes d'apprentissage	36
figure 9	Influence du nombre d'attributs de textures.....	37
figure 10	Technique de localisation des marqueurs.....	39
figure 11	Segmentation des couronnes des cocotiers	39
figure 12	Échantillonnage terrain prévu et effectué.....	40
figure 13	Strates échantillonnées des cocotiers	40
figure 14	3 types de cocoteraies.....	41
figure 15	Classification SVM de la cocoteraie du motu principal de Tikehau.....	42
figure 16	Processus ponctuels de Poisson homogène et non-homogène	42
figure 17	Estimation de λ pour une sous-région du nuage de points réel.....	43
figure 18	Histogramme des densités λ et segmentation de la cocoteraie.....	43
figure 19	Exemples des différentes classes.....	44
figure 20	Méthode de classification par fouille de données	46
figure 21	Composite AirSAR (C_{VV} , L_{HV} , L_{HH}) - Moorea.....	47
figure 22	Composite AirSAR (C_{VV} , L_{HV} , L_{HH}) - domaine d'Opunohu	48
figure 23	Signature normalisée du jeu de classes	49
figure 24	Classification max. de vraisemblance AirSAR – Opunohu	50
figure 25	Indices NDVI (gauche) et SARvi (droite) - Opunohu	52
figure 26	Précision des classifications SVM de triplet de bandes radar	53
figure 27	Composite JPL-AirSAR (C_{VV} , L_{HV} , L_{HH}) Tubuai.....	54
figure 28	Classification de Tubuai.....	56
figure 29	Contribution des indices polarimétriques par bandes.	60
figure 30	Cartographie SVM de l'habitat potentiel du Miconia – Moorea	64
figure 31	Zones d'études - Moorea.....	69
figure 32	Précision globale (OA) comparative sur les ATE.....	70
figure 33	Précision globale (OA) comparative sur les ME.....	70
figure 34	Précision globale (OA) comparative sur les NTE.....	70

figure 35	Concept de l'algorithme SVM.	71
figure 36	Taux de bonne reconnaissance en fonction du nombre d'itérations	73
figure 37	Optimisation systématique vs optimisation unique.....	74
figure 38	Importance de la ré-optimisation des paramètres SVM.....	75
figure 39	Classification radar SVM - Tubuai	77
figure 40	Emplacement des écotones.....	78
figure 41	Méthodologie de repérage et d'échantillonnage des écotones	79
figure 42	Comparaison croisée algorithme / échantillonnage terrain	79
figure 43	Classification SVM avec échantillonnages sur les écotones– Mont Marau.....	81
figure 44	Principe de la fusion SVM	84
figure 45	Précision par classe des classifications mono-sources et fusion SVM	86
figure 46	Classification par fusion SVM multi-sources - Moorea.....	87
figure 47	Localisation des échantillonnages terrain - Moorea.....	88
figure 48	Principe de la fusion sélective	89
figure 49	Précision globale en fonction du seuil α	92
figure 50	Carte des milieux naturels de Moorea obtenue par fusion SVM sélective	93
figure 51	Chaîne de traitement pour la cartographie du Miconia	94
figure 52	Cartographie SVM du Miconia - vallée de la Papenoo - Tahiti.....	95
figure 53	Schéma de la comparaison des algorithmes Random Forest et SVM.....	96
figure 54	Carte de végétation de Moorea par classification SVM d'image Quickbird	97
figure 55	Espèces rares : a) <i>Lepinia taitensis</i> , b) <i>Pouteria tahitensis</i> , c) <i>Santalum insulare</i>	97
figure 56	Échantillonnages terrains	98
figure 57	Cartographie RF et SVM de l'habitat potentiel de 3 espèces rares - Moorea	98
figure 58	Bassins versants de Moorea	99
figure 59	Variables sociologiques de Moorea	100
figure 60	Variables physiographiques de Moorea	101
figure 61	Influence relative de chaque variable sur la santé du corail (RHI)	102
figure 62	Image Keyhole7 de l'atoll de Mururoa (1967).....	106
figure 63	Image Worlview2 de Raiatea et Tahaa (2011).....	116

LISTE DES ACRONYMES

ACP	Analyse en composante principale
AGU	American Geophysical Union
AIRSAR	Airborn Synthetic Aperture Radar
ALOS	Advanced Land Observing Satellite
ANSI	American National Standards Institute
AP	Alternate Polarization
API	Annual Potential Insolation
AQUARIUS	Satellite mesurant la salinité de la surface de la mer
ASAR	Advanced Synthetic Aperture Radar
ASTER	Advanced Spaceborn Thermal Emission and Reflection radiometer
ATE	Anthropogenic Terrestrial Ecosystems
BQR	Bonus Qualité Recherche
CEP	Centre d'Expérimentation du Pacifique
CNES	Centre National d'Etudes Spatiales
CNRS	Centre National de la Recherche Scientifique
CNU	Conseil National des Universités
CRCT	Congé pour Recherche ou Reconversion Thématique
CTI	Compound Topographic Index
DEA	Diplôme d'Etudes Approfondies
DEM	Digital Elevation Model
DLR	German Aerospace Center
DP	Dual Polarimetry
DUT	Diplôme Universitaire de Technologie
EA	Equipe d'Accueil
EEA	Electricité, Electronique, Automatique
ENSG	Ecole Nationale des Sciences Géographiques
ENVISAT	ENVironment SATellite
ESEO	Ecole Supérieur d'Electronique de l'Ouest
FP	Full Polarization
GARP	Genetic Algorithm for Rule set Production
GDR	Groupement de Recherche
GGI	Géo-matériaux et Géologie de l'Ingénieur
GLCM	Gray-Level Co-occurrence Matrices
GMT	Greenwich Mean Time
GPS	Global Positioning System
GRP	Good recognition percentage
HDR	Habilitation à Diriger des Recherches
HR	Haute Résolution
HTML	Hypertext Markup Language
IEEE	Institute of Electrical and Electronics Engineers
IGN	Institut Géographique National
INPG	Institut National Polytechnique de Grenoble
INSA	Institut National des Sciences Appliquées
INSU	Institut National des Sciences de l'Univers
IR	Infrarouge
IRD	Institut de Recherche pour le Développement
ISO	International Standard Organization
JPL	Jet Propulsion Laboratory

KH	Key Hole
LEA	Langues Etrangères Appliquées
MAJECSTIC	MANifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication
MCF	Maître de ConFérences
ME	Marine Ecosystems
ML	Maximum Likelihood
MLC	Multi-Look Complex
MNE	Modèle Numérique d'Elevation
MODIS	Moderate-Resolution Imaging Spectroradiometer
MPA	Mean Producer Accuracy
NASA	National Aeronautics and Space Administration
NDVI	Normalized Difference Vegetation Index
NTE	Natural Terrestrial Ecosystems
OA	Overall Accuracy
ORSTOM	Office de la Recherche Scientifique et Technique Outre-Mer
OTB	Orpheo Tool-Box
PACRIM	Pacific Rim
PALSAR	Phased Array L-band Synthetic Aperture Radar
PCA	Principal Component Analysis
PF	Polynésie Française
PPF	Plan Pluri-Formation
PPME	Pôle Pluridisciplinaire de la Matière et de l'Environnement
RBF	Radial Basis Function
RF	Random Forest
RGB	Red Green Blue
RHI	Reef Health Index
ROI	Region Of Interest
RVB	Rouge Vert Bleu
SAR	Synthetic Aperture Radar
SAU	Service de l'Urbanisme
SDR	Service du Développement Rural
SELF	Selective Fusion
SEOM	Secrétariat d'Etat à l'Outre-Mer
SIG	Système d'Information Géographique
SPIE	International Society for Optics and Photonics
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SVM	Support Vector Machine
TD	Travaux dirigés
THR	Très Haute Résolution
UNC	Université de la Nouvelle-Calédonie
UPF	Université de la Polynésie Française
UPMC	Université Pierre et Marie Curie
USA	United States of America
ZEE	Zone Economique Exclusive

PREAMBULE

”La **télétection** est la technique qui, par l’acquisition d’images, permet d’obtenir de l’information sur la surface de la terre sans contact direct avec celle-ci. La télétection englobe tout le processus qui consiste à capter et à enregistrer l’énergie d’un rayonnement électromagnétique émis ou réfléchi, à traiter et à analyser l’information, pour ensuite mettre en application cette information.”. (Centre canadien de télétection <http://www.rncan.gc.ca>)

Un des aspects de la télétection est donc d’extraire des informations d’images (satellites ou aériennes) et d’en restituer une représentation synthétique et intelligible. En ce sens, une analogie entre la télétection et la peinture est envisageable (*tableau 1*).

Peintre	↔	Télédetecteur
Sujet	↔	Zone d’étude
Vision du peintre	↔	Vérité terrain
Couleurs	↔	Jeu de classes
Support	↔	Image
Préparation du support	↔	Prétraitement
Pinceaux	↔	Algorithmes
Techniques picturales	↔	Méthodologies
Œuvre	↔	Carte thématique

tableau 1 Analogie entre la peinture et la télétection

Un peintre cherche à représenter un sujet qu’il choisit pour son intérêt esthétique et/ou métaphorique, le télédetecteur cherche quant à lui à représenter une zone d’étude ayant un intérêt par rapport à une problématique donnée (environnementale, sociétale, ...).

Le travail de l’artiste est contraint par son contexte social et son contexte géographique. Le contexte social décrit les conditions matérielles auquel est soumis le peintre, a-t-il accès à un atelier ? A quelles sources de financements peut-il prétendre ? Dans quel milieu évolue-t-il ? Fait-il partie d’une communauté ou d’un réseau d’artistes ? Tous ces éléments vont engendrer des contraintes ou des libertés sur sa capacité à travailler et à produire des œuvres. Le contexte géographique va aussi influencer le choix du sujet de l’œuvre, un peintre habitant la Bretagne ne choisira pas les mêmes sujets qu’un peintre résidant aux îles Marquises (*cf.* Paul Gauguin). Ainsi, le contexte (social et géographique) va être un facteur déterminant du déroulement de la carrière d’un enseignant-chercheur et des axes de recherches qu’il va choisir de développer.

La vision du peintre est essentielle dans tout travail artistique. Elle est une étape préliminaire permettant de rassembler les éléments que le peintre va chercher à représenter, son point de vue sur le sujet, ce qui lui semble le plus important et pertinent pour obtenir la représentation qu’il souhaite. Il en est de même pour le télédetecteur qui, par le biais de la vérité terrain, va analyser et choisir les éléments qu’il souhaite détecter et mettre en valeur dans un paysage. Comme pour le peintre, cette étape est primordiale et va conditionner les résultats obtenus *in fine*.

Le peintre va ensuite opérer un choix sur les couleurs et matières qu’il va utiliser : ce choix dépend de sa vision du sujet, et l’aspect final de l’œuvre en sera donc affecté. Le télédetecteur choisit son jeu de classes, c’est-à-dire une liste synthétique et limitée des éléments les plus importants du paysage qu’il va tenter de détecter et qui composeront sa carte finale.

L’artiste choisit le support de son œuvre (par exemple : toile, papier, parchemin, peau, ...). Ce choix aura une influence sur le rendu final mais surtout imposera l’utilisation de certains

outils, certaines matières et certaines techniques picturales. De son côté, le télédéetecteur choisit une ou plusieurs images sources, c'est-à-dire la donnée qui va servir de base à ses traitements et ce choix devra être pris en considération lors du choix des algorithmes et méthodologies utilisés.

Un peintre doit dans certains cas préparer le support. Cette étape, peu connue, est d'une importance capitale sur le rendu final (la formation des peintres sur cet aspect pouvait prendre plusieurs années). Il en va de même pour le télédéetecteur qui doit (faire) effectuer des prétraitements sur les images : corrections géométriques, radiométriques, géolocalisation, etc.

Le peintre dispose d'un certain nombre d'outils comme les pinceaux, couteaux, crayons, fusains, etc. Le télédéetecteur dispose d'algorithmes, il en existe de nombreux, dans tous les domaines. Leur étude et comparaison est un point de passage obligé pour en comprendre le fonctionnement et ainsi pouvoir les utiliser à bon escient.

Une fois effectués tous ces choix, le peintre applique des techniques picturales qui lui permettent d'obtenir son œuvre finale. Ces techniques picturales ont un lien direct avec le type d'œuvre qu'il souhaite produire. De la même façon, le télédéetecteur met en œuvre une méthodologie ou suite de traitements qui, à partir des éléments précédemment choisis, lui permettent d'obtenir le résultat escompté.

La *figure 1* est une suite de lithographies de Picasso sur le thème du taureau, elle montre bien qu'à partir d'un sujet unique, les choix successifs d'un artiste vont donner des résultats très différents. De la même façon, les choix techniques successifs d'un télédéetecteur vont avoir une importance capitale sur la cartographie finale.

« S'il n'existait qu'une seule vérité, on ne pourrait peindre des centaines de tableaux sur un même sujet. » Pablo Picasso

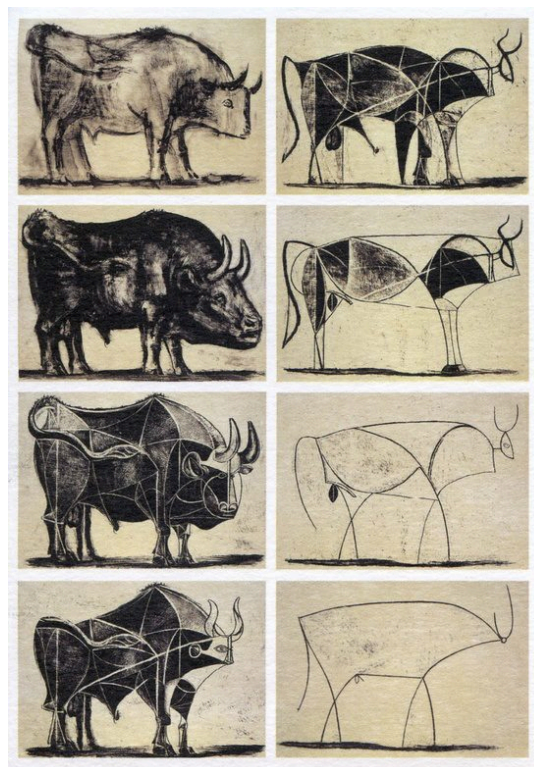


figure 1 Le taureau de Picasso (1945)

I. INTRODUCTION

J'ai été recruté en août 2001 comme maître de conférences au sein du laboratoire Terre-Océan¹ de l'université de la Polynésie française pour mon profil 61^{ème} section du CNU : Traitement du signal et des images.

Mes travaux de thèse portaient sur la séparation de sources, branche du traitement de signal numérique. Continuer ces travaux, fortement théoriques, au sein d'une université ultra-marine n'aurait pas eu beaucoup de sens. Je me suis donc engagé à effectuer une reconversion vers une thématique de recherche plus appliquée et en particulier vers le traitement d'images de télédétection locales.

En l'an 2000, juste avant mon arrivée au sein du laboratoire, la campagne PACRIM-II du Jet Propulsion Lab. de la NASA a effectué l'acquisition d'images sur les principales îles de la Polynésie française à l'aide de deux capteurs aéroportés : AirSAR (radar polarimétrique) et Modis/ASTER (optique hyper-spectral). Les données radar polarimétriques AirSAR ont fait l'objet de ma première étude. Cependant, aussi exceptionnel que soit ce jeu de données, il ne pouvait constituer un axe de recherche en soi.

A. AXE DE RECHERCHE

Je me suis donc orienté vers la thématique plus large de la « **télédétection satellitaire en milieu insulaire polynésien** ». L'utilisation de la télédétection s'y justifie pleinement par l'accès difficile lié à l'isolement géographique des îles et aux reliefs escarpés qui rendent couteuse et compliquée toute tentative de travail sur le terrain.

Les îles polynésiennes sont un écosystème à part entière, du sommet de la montagne au pied du volcan sous-marin qui lui a donné naissance en passant par le récif corallien et son lagon. Cet écosystème est néanmoins naturellement divisé en une partie marine d'une part et une partie terrestre d'autre part.

Le milieu marin récolte sans conteste le plus d'attraits en Polynésie avec son récif corallien important, mais aussi ses ressources halieutiques. Il fait l'objet de nombreuses études scientifiques dans tous les domaines depuis de nombreuses années.

En revanche, le milieu terrestre, bien que moins médiatisé et mis en valeur, n'en demeure pas moins très particulier et très intéressant. Sa partie naturelle contient l'intégralité de la biodiversité végétale polynésienne marquée par un fort taux d'endémisme et sa partie anthropisée, bien que peu étendue, est le siège de nombreux enjeux environnementaux et sociétaux.

Ce milieu terrestre est une des grandes richesses de la Polynésie française et doit, au même titre que le milieu marin, être étudié, géré et protégé. Nous pouvons aussi considérer que la volonté de préserver le récif corallien et les ressources marines sans prendre en compte la partie immergée de l'île, sa couverture naturelle mais aussi ses zones anthropisées n'a pas de sens.

C'est la raison pour laquelle j'ai choisi de travailler sur le milieu naturel terrestre en complément du travail des spécialistes de ces écosystèmes. Je me suis donc consacré, au cours de ces douze dernières années, à l'étude de la répartition de la végétation des îles polynésiennes par télédétection satellitaire.

¹ renommé depuis Géopôle du Pacifique Sud - GePaSUD – EA4238

B. CONTEXTE SOCIAL

Le contexte social d'un enseignant-chercheur est un élément important pour comprendre ses conditions de travail et explique souvent l'orientation de sa carrière. En Polynésie française, celui-ci diffère fortement de celui auquel j'ai été confronté en métropole, et c'est en toute connaissance de cause que j'ai accepté ce poste de maître de conférences à l'université de la Polynésie française.

1. Facteurs contraignants

A mon arrivée, le laboratoire Terre-Océan était un laboratoire éminemment multidisciplinaire, avec un seul chercheur par spécialité : biologie marine spécialiste des algues, informaticien, physique de l'atmosphère, biologie animale spécialiste des cétacés, courantologie océanique, géophysique du globe, etc.

Faire partie d'un laboratoire multidisciplinaire constitue un avantage indéniable en ce sens que cela offre l'opportunité d'acquérir une ouverture scientifique des plus enrichissantes. Toutefois, ce choix m'exposait à un isolement scientifique significatif. En effet, j'étais le seul chercheur en traitement du signal et des images au sein de mon laboratoire, de mon université et même de Polynésie, et je ne pouvais bénéficier sur place d'une émulation scientifique, saine et nécessaire.

Pour pallier ce handicap, je me suis déplacé autant que possible en métropole et dans les conférences internationales afin de pouvoir présenter mes résultats de recherche, recueillir l'avis de mes collègues, me tenir au fait des dernières techniques, mais surtout créer et entretenir un réseau de collaborations.

Je suis membre IEEE depuis 2002 dans les communautés IEEE Remote Sensing et Signal Processing.

A cet isolement déjà contraignant s'ajoute le problème de l'éloignement: 20 000km de distance avec la métropole et mes collègues métropolitains, un trajet de 20h par avion dont le billet est facturé entre 2 et 3 000 euros par personne pour un vol sec, 12h de décalage horaire. Tous ces éléments logistiques compliquent fortement l'organisation de missions de recherche.

De fait, je devais conduire ma recherche en autonomie, c'est-à-dire créer ma propre dynamique scientifique, définir mes axes de recherche, organiser mes études scientifiques, rechercher des financements, sélectionner et encadrer des stagiaires et enfin publier.

Le métier d'enseignant-chercheur comporte trois activités d'égale importance : la recherche, l'enseignement et l'administration. Or, j'étais conscient de l'importance de l'activité de recherche pour l'évolution de carrière et aussi qu'il me faudrait redoubler d'efforts pour conserver une activité de recherche régulière et productive tout en assumant mes responsabilités d'enseignement et d'administration.

2. Enseignement

L'université de la Polynésie française est une petite université de 3000 étudiants et une centaine d'enseignants permanents. Cela implique que l'on assume une charge d'enseignement conséquente. J'ai accepté de façon régulière un sur-service de 14 à 72%, ce qui correspond à un volume annuel moyen de 260h équivalents TD (*tableau 2*).

Mes enseignements se font exclusivement en informatique puisque mon poste était profilé 27/61 : Informatique pour l'enseignement et traitement du signal et des images pour la recherche.

Enseignement	Département	Niveau	Années														
			2001-2002	2002-2003	2003-2004	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013			
Informatique générale	Sciences	L1															
Techniques informatiques	Sciences	L1															
Introduction aux réseaux	Sciences	L1															
Algorithmique	Sciences	L1															
Algo. et prog. C niveau 1	Sciences	L1															
Bureautique	Lettres	L2															
C2i niveau 1	Lettres	L2															
HTML/Javascript	Sciences	L2															
HTML/Css	Sciences	L2															
Javascript	Sciences	L2															
Algo. et prog. C niveau 2	Sciences	L2															
Algo. et prog. C Avancée	Sciences	L2 - L3															
Edition de pages web avancée	Sciences	L3															
Calcul scientifique	Sciences	L3															

Volume eq. TD 236h 224h 242h 285h 258h 258h 220h 235h 251h 98h 312h 330h

tableau 2 Tableau des enseignements

3. Responsabilités

J'ai été élu au sein de la commission de spécialistes n°4 de l'UPF, ce qui m'a permis de prendre part aux recrutements de plusieurs de mes collègues.

J'ai aussi été par trois fois élu comme membre du conseil scientifique de l'UPF, contribuant ainsi à la politique de recherche de l'UPF.

J'ai été responsable pédagogique, mais surtout responsable de la mise en place du Certificat Informatique et Internet à l'UPF, j'ai tout d'abord commencé par organiser le C2i niveau 1 dans les filières d'enseignements, géré la certification de nos étudiants puis j'ai organisé la mise en place du C2i2e (C2i niveau 2 Enseignant) dont je reste le responsable et le président de jury de certification.

Enfin, j'ai été nommé par deux fois président de jury de Bac.

Mandats électoraux

De 2003 à 2008	Membre de la commission de spécialiste n°4 de l'UPF.
De 2007 à 2013	Membre élu (collège des enseignants non HDR) du conseil scientifique de l'UPF.

Chargé de mission

Depuis 2005	Correspondant C2I (Certificat informatique et internet niveau 1 et niveau 2 enseignant). Chargé de mission pour la mise en place du C2i niveau 1. Chargé de mission pour la mise en place du C2i niveau 2E.
-------------	---

Responsabilités pédagogiques

Depuis 2001	Coordinateur de discipline en informatique.
2007-2008 et 2012-2013	Responsabilité pédagogique L2 licence mathématiques et informatique.

Participation à des réseaux de chercheurs

Depuis 2000	Membre du GDR-Isis.
Depuis 2002	Membre IEEE et membre du Web INSU CNRS.
Depuis 2003	Membre du réseau chercheur télédétection de l'AUF.

J'ai toujours veillé à ce que ma charge d'enseignement et responsabilités ne soient jamais préjudiciables à mon activité de recherche, comme l'atteste la concrétisation de ce mémoire d'HDR

4. Recherche

a. Encadrements

Pour pouvoir mener à bien une activité de recherche tout en assumant les différentes responsabilités administratives et d'enseignement, il faut encadrer des stagiaires de tous niveaux, et pour cela avoir accès à un vivier d'étudiants pour effectuer des stages ou thèses.

Malheureusement, l'UPF ne forme que jusqu'au niveau licence en sciences, je me suis donc rapproché des universités métropolitaines pour recruter des stagiaires. Je souhaite préciser ici

que tous les collègues que j'ai pu contacter ont fait preuve d'une grande solidarité en relayant mes offres.

En outre, proposer un stage en Polynésie implique de trouver des moyens importants pour financer le déplacement, le logement et le défraiement des étudiants (il faut compter au moins 3 000 euros pour un stage de 3 mois) ce qui, constitue une évidente restriction.

Thèses

- 2008 Co-encadrement de thèse de **Cédric Lardeux**, Université de Paris-Est, Marnes-La-Vallée, *Apport de la polarimétrie radar à la cartographie thématique du couvert végétal*, directeur de thèse : Jean-Paul Rudant, co-encadrants : Pierre-Louis Frison, Jean-Claude Souyris, Benoît Stoll, date de soutenance : 09 décembre 2008, (*Lardeux thesis2008*).
- 2009 Co-direction de thèse de **Raimana Teina**, Université Pierre et Marie Curie, *Caractérisation de la cocoteraie des Tuamotu à partir d'images satellites à très haute résolution spatiale*, directeur de thèse : René Alt (UPMC) et Benoît Stoll (UPF), co-encadrant : Dominique Béréziat (UPMC), date de soutenance : 15 septembre 2009, (*Teina thesis2009*).
- 2011 Co-encadrement de thèse de **Robin Pouteau**, *Étude de la phytogéographie des îles hautes de Polynésie française par classification SVM d'images multi-sources*, directeur de Thèse : Alban Gabillon (UPF), co-encadrant : Benoît Stoll (UPF), date de soutenance : 07 novembre 2011, (*Pouteau thesis2011*).

Stages

- 2003 Encadrement de stage de **Johann Thomas**, ENSG-IGN IT2, *Étude de la cartographie du couvert végétal des îles hautes polynésiennes*, durée 3 mois, (*Thomas report2003*).
- 2004 Encadrement de stage de **Rémi Provin**, ENSG-IGN IT2, *Cartographie de la végétation des îles australes à partir d'images AirSAR ou MASTER*, durée 3 mois, (*Provin report2004*).
- 2005 Co-encadrement de stage **Cédric Lardeux**, ENSG-IGN, M2 Sciences de l'information géographique, *Apport de la polarimétrie radar à la cartographie thématique en Polynésie française*, durée 6 mois, (*Lardeux report2005*).
- 2009 Co-encadrement de stage **Manuel Castela**, INPG (ENSE3) M2 ATIS, *Étude comparative d'attributs et descripteur de texture et application à la classification d'images optiques haute résolution du plateau de Taravao - TAHITI*, durée 6 mois, (*Castela report2009*).
- 2010 Co-encadrement de stage **Élise Desmier**, INSA Toulouse, *Étude de la cocoteraie des Tuamotu par fouille de données*, durée 8 mois, (*Desmier report2010*).
- 2010 Encadrement de stage de **Romuald Ballet**, ENSIETA Brest, *Cartographie de la végétation de l'île de Tahiti dans un contexte d'érosion sur données SPOT5*, durée 2 mois, (*Ballet report2010*).
- 2011 Co-encadrement de **Jean-Baptiste Goujeon**, groupe ESEO: Anger, *Optimisation de la classification de donnée multi-spectrale satellitaire THR, application sur les îles hautes polynésiennes*, durée 3 mois, (*Goujeon report2011*).

Année	Encadrements		Publications				
	Stages	thèses	Posters	Conférences	Articles		
						grsl	jstars
2001							
2002	J. Thomas						
2003	R. Provin		<i>agu</i>	<i>igarss</i>			
2004	C. Lardeux		<i>arfp</i>	<i>igarss</i>			
2005				<i>igarss</i>			
2006		R. Teina	<i>perf</i>	<i>igarss</i>	<i>majestic</i>		
2007		C. Lardeux			<i>envisat</i>		
2008	M. Castela		<i>psi</i>	<i>igarss</i>	<i>polinsar</i>	<i>igarss</i>	
2009	E. Desmier	R. Pouteau	<i>alos</i>	<i>spie</i>	<i>alos</i>	<i>spie</i>	
2009	R. Ballet					<i>spie</i>	
2010	J.-B. Goujeon				<i>ipta</i>	<i>igarss</i>	
2011				<i>igarss</i>		<i>icdim</i>	
2012					<i>aprs</i>	<i>isrse</i>	
2013				<i>igarss</i>		<i>isrse</i>	

tableau 3 Encadrements et publications scientifiques

b. Financements

Cela nous conduit à la problématique de la recherche de financements. La Polynésie étant tournée naturellement vers le tourisme et les produits de la pêche et de la perliculture, les fonds publics auxquels on peut avoir accès sur une thématique environnement et végétation sont très limités. J'ai malgré tout été en mesure de trouver des sources de financements auprès de services comme la délégation à la recherche, le service du développement rural entre autres.

Contrats de recherche

- 2006 Service du développement rural, *étude pilote de la cocoteraie des Tuamotu sur des images Ikonos (Tikehau)* (Teina report2007), montant 4 200 €
- 2006 BQR université de la Polynésie française; *achat de deux images satellites haute résolution SPOT5 de l'île de Tahiti*, programme ISIS (SpotImage) montant 1 200 €
- 2009 Délégation à la recherche / Biocode, *étude de la cartographie du milieu naturel de Moorea par classification SVM Conjointe de données satellites radar et optique à très haute résolution* (Pouteau et al. report2009), montant 12 500 €
- 2011 Obtention d'un congé pour recherche ou conversion thématique (CRCT) de 6 mois entre janvier et juin 2011.

Plusieurs autres financements ponctuels de moindre importance ont été obtenus par le biais du bonus qualité recherche (BQR) de l'UPF.

Financements de stages

- 2003 Stage **Johann Thomas** (billet avion + *Per diems*) financé sur crédits SEOM (secrétariat d'état à l'outre-mer) «Cartographie et imagerie radar et multi-spectrale de la Polynésie française» (responsable Patrick Capolsini - UPF).
- 2004 Stage **Rémi Provin** (billet avion) financé par la délégation à la recherche de la Polynésie française.
- 2005 Stage **Cédric Lardeux** financé par l'ENSG Marnes-la-Vallée.
- 2009 Stage **Manuel Castela** (billet avion) financé par l'IRD Nouméa - US espace.
- 2010 Stage **Romuald Ballet** financé par l'ENSIETA Brest.
- 2010 Stage **Élise Desmier** financé par le PPF Nouméa «Ingénierie de connaissances pour l'aide à la gestion intégrée des littoraux insulaires tropicaux», UNC.

Acquisition d'images satellites

- 2007 Couple de scènes SPOT5 de Tahiti (30 mai 2002 et 6 mai 2009), programme de financement ISIS de SpotImage, montant 1 200 €
- 2009 Scène Quickbird de Moorea (9 novembre2006), montant 2 500 €
- 2009 Scènes TerraSAR-X de Moorea (8 scènes entre le 29 avril et le 5 novembre 2010), programme d'aide à la recherche du DLR, montant 1 000 €

Cela dit, travailler en Polynésie française ne comporte pas que des contraintes, cela présente aussi des avantages. L'élément qui m'a sans conteste été le plus utile est le fait que depuis une dizaine d'années le territoire de la Polynésie française a fait l'acquisition de données satellitaires pour ses besoins cartographiques, son cadastre informatisé et la production de modèles numériques de terrain.

Une convention signée avec le service de l'urbanisme du territoire de la Polynésie française permet l'accès à son fonds cartographique pour tous les chercheurs de l'UPF. Avoir accès gratuitement à des données satellites de qualité constitue un véritable atout, alors qu'il s'agit souvent d'un poste budgétaire très lourd des études en télédétection.

c. Collaborations

Le contexte géographique polynésien en général pose de nombreuses questions scientifiques, et la thématique terrestre en particulier ouvre sur de nombreuses problématiques environnementales : préservation de la biodiversité, problématique des espèces envahissantes, impact anthropique sur le milieu naturel, ... Cela suscite un intérêt marqué des équipes métropolitaines et internationales pour travailler sur les problématiques et données locales nous offrant alors l'opportunité de concrétiser des collaborations intéressantes pour les deux parties, comme en témoigne la constitution récente de réseaux de recherche comme le GOPS (Grand Observatoire du Pacifique Sud), le réseau PACENET (Pacific Europe Network) entre autres.

J'ai ainsi pu collaborer avec différents laboratoires comme le laboratoire d'informatique de Paris 6 (LIP6) de l'université Pierre et Marie Curie, le G2I-IFSA et l'Institut Francilien des Géosciences de l'université de Marnes-La-Vallée, le DCT/SI/AR du centre national d'études spatiales (CNES) de Toulouse, le LIRIS Combining de l'INSA-Lyon, le laboratoire PPME de l'université de la Nouvelle-Calédonie et l'antenne de l'IRD à Nouméa, Nouvelle-Calédonie.

C. CONTEXTE GEOGRAPHIQUE

Le contexte géographique singulier de la Polynésie française est un facteur qui influence fortement mes choix scientifiques. Prenons le temps d'en exposer ici les principales caractéristiques.

1. Genèse des îles volcaniques

La Polynésie française est un territoire essentiellement maritime, vaste comme l'Europe. Elle est composée d'à peu près 120 îles d'une surface totale cumulée de 4 000 km² réparties sur une zone économique exclusive de 5 millions de km².

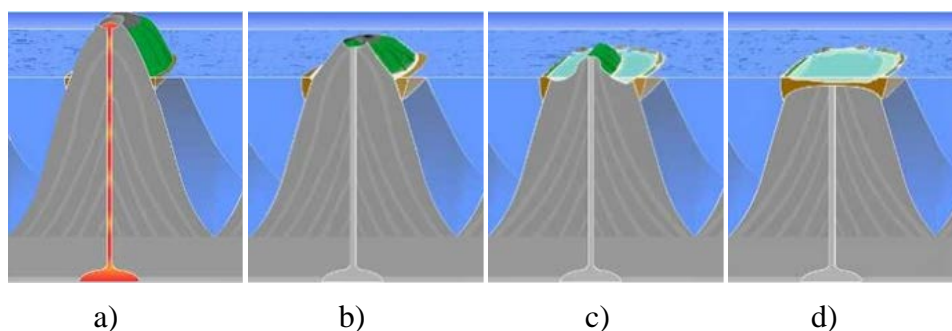


figure 2 Évolution des îles coralliennes²

Il existe en Polynésie française plusieurs types d'îles toutes d'origine volcanique : au départ, un volcan sous-marin apparaît au-dessus d'un point chaud du plancher océanique, celui-ci finit par émerger et créer une île volcanique (figure 2-a). Du corail se développe sur son littoral, il s'agit du récif frangeant (figure 2-b). C'est la configuration que l'on retrouve aux îles Marquises par exemple.

² <http://coraux.univ-reunion.fr/>

Une fois refroidie, l'île s'enfonce lentement, sous l'effet de son poids, dans le plancher océanique. Le corail s'enfonce mécaniquement avec l'île et croît vers la surface pour continuer à bénéficier d'un ensoleillement optimal. Sous l'effet des courants, un chenal apparaît, scindant le récif corallien en un récif frangeant et un récif barrière (*figure 2-c*). C'est la configuration de la plupart des îles de l'archipel de la Société comme Tahiti, Moorea, Raiatea etc.

Enfin, lorsque l'île principale a complètement disparu par subsidence, il ne reste que le récif barrière, c'est devenu un atoll (*figure 2-d*), dernier stade de l'évolution des îles polynésiennes. Cela correspond à la configuration des atolls des Tuamotu.

Par soucis de clarté, nous ne parlerons dans ce document que de deux types d'îles : les îles hautes et les atolls qui correspondent respectivement au début et à la fin de l'évolution des îles volcaniques tropicales et qui possèdent deux types bien distincts de couvertures végétales.

2. Répartition spatiale de la végétation

Le *tableau 4* illustre la faible taille des principales îles hautes polynésiennes avec une centaine de kilomètres carrés (à l'exception de Tahiti qui fait à peu près 1 000 km²) et leur topographie escarpée avec des sommets culminants dans les 1 000 mètres (2 241 mètres pour Tahiti). Ces caractéristiques ont une incidence forte sur l'organisation spatiale des paysages observés et la répartition des différents types de couverts végétaux.

île	Superficie	Altitude max	population (2007)	Archipel
Tahiti	1052 km ²	2 241 m	176 000	Société
Ua Pou	105 km ²	1 230 m	2 200	Marquises
Nuku Hiva	387 km ²	1 224 m	2 700	Marquises
Hiva Oa	326 km ²	1 213 m	2 000	Marquises
Moorea	134 km ²	1 207 m	16 000	Société
Tauata	61 km ²	1 050 m	700	Marquises
Ra'iatea	238 km ²	1 017 m	12 000	Société
Fatu Hiva	84 km ²	960 m	600	Marquises
Ua Huka	84 km ²	857 m	600	Marquises
Bora Bora	40 km ²	727 m	9 000	Société
Huahine	75 km ²	669 m	6 000	Société
Rapa	40 km ²	650 m	500	Australes
Taha'a	90 km ²	590 m	5 000	Société
Raivavae	20 km ²	437 m	900	Australes
Tubuai	45 km ²	422 m	2 000	Australes
Rurutu	33 km ²	412 m	2 100	Australes
Maupiti	11 km ²	380 m	1 200	Société
Rimatara	9 km ²	83 m	800	Australes

tableau 4 Principales îles hautes de Polynésie française³

a. Phytogéographie

La phytogéographie étudie l'action du milieu sur la répartition des plantes. Selon celle-ci, la répartition spatiale ou horizontale (que je qualifierai parfois de spatialisation) des espèces végétales serait fortement liée à un certain nombre de facteurs géophysiques répartis en deux groupes, les facteurs abiotiques et biotiques :

³ ISPF (Institut de la Statistique de Polynésie française), "Atlas démographique de la Polynésie française : recensement de la population 2007", *Regards special issue 2*, 2007.

Le premier groupe est constitué des facteurs abiotiques, c'est-à-dire les facteurs environnementaux non-vivants influant directement sur les conditions de vie des espèces végétales. Ils sont de plusieurs types :

- les facteurs édaphiques (du sol) : structure du sol, sels minéraux, ...
- les facteurs climatiques : température, lumière, eau, ...
- les facteurs chimiques : concentration en gaz dans l'atmosphère, ...
- les facteurs topographiques : altitude, pente, ...

Ces facteurs abiotiques sont relativement aisés à modéliser, soit à partir d'un modèle numérique d'élévation, soit à partir de carte de précipitations par exemple.

Le second groupe est constitué des facteurs biotiques qui représentent les actions du vivant sur le vivant comme les ressources alimentaires, les relations trophiques de prédation, coopération, compétition, parasitisme, etc. Ces facteurs biotiques sont plus compliqués à modéliser.

b. Effet de foehn

A cela, il faut rajouter l'influence des vents dominants sur l'étagement de la végétation. Ce phénomène bien connu dans les paysages montagneux continentaux est le fruit de la conjonction des vents dominants avec une chaîne de montagne qui engendre une différenciation du climat entre le côté exposé aux vents et celui à l'abri du vent, il s'agit de l'effet de foehn.

Ce phénomène prend une dimension particulière sur les îles polynésiennes, la topographie très accentuée (sommets escarpés et fortes pentes) et des vents dominants importants (les fameux alizés) engendrent une différenciation très marquée du climat (topoclimat), et donc de la spatialisation de la végétation, en fonction de la côte considérée : humide sur la côte sud-est et sec sur la côte nord-ouest.

La *figure 3* illustre l'altitude des différents types de végétation selon un axe nord-ouest, sud-est de l'île de Tahiti.

La conjonction de tous ces phénomènes engendre donc une spatialisation des espèces naturellement hétérogène et complexe qui est la caractéristique principale de la couverture végétale des îles hautes polynésiennes.

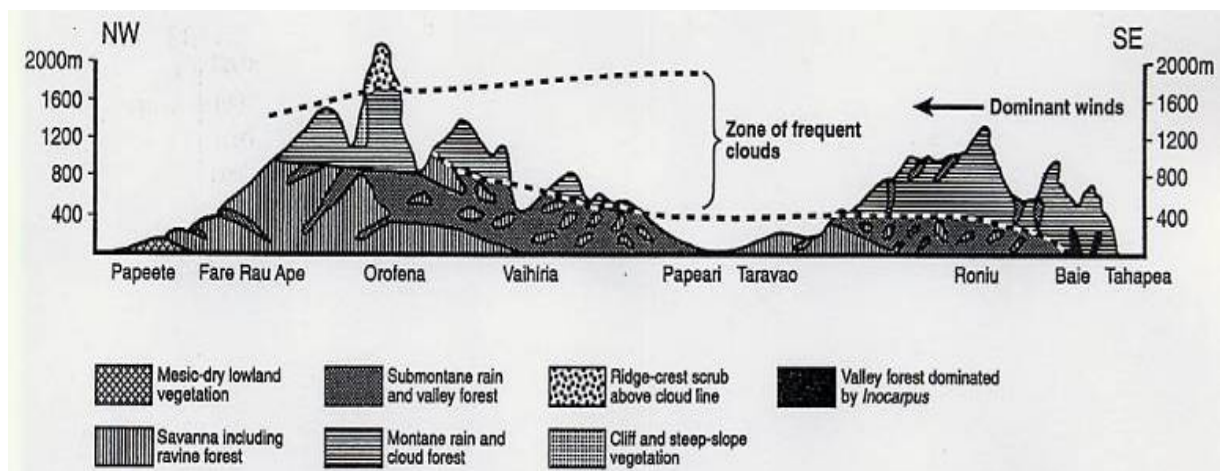


figure 3 Influence des vents dominants - Tahiti⁴

⁴ Papy, 1951-1954 in Mueller-Dombois et Fosberg, 1998

c. Structure verticale de la végétation

Un dernier degré de complexité est engendré par la structure verticale des forêts tropicales. Celle-ci peut comporter quatre ou cinq strates selon les sources : schématiquement, il existe une strate muscinale (mousses et champignons), une strate herbacée, une strate arbustive et une strate arborescente.

La télédétection satellitaire directe ne concerne en général que la strate supérieure de la végétation qu'elle soit herbacée, arbustive, arborescente ou mixte : on ne peut en effet détecter que ce que l'on « voit » sur une image. Celle-ci étant une vue aérienne, on ne perçoit que la partie la plus élevée de la végétation et rarement les strates inférieures.

Nous verrons que dans certains cas, nous pouvons accéder aux étages inférieurs de la végétation en utilisant des méthodes de télédétection indirectes.

d. Facteurs anthropiques

Les règles écologiques précédemment décrites ont créé de façon naturelle un certain nombre de paysages typiques des îles hautes Polynésiennes et la dynamique de la végétation était certainement arrivée à un relatif équilibre avant que l'influence anthropique ne s'accroisse.

La *figure 4* illustre un paysage de Tahiti en 1960 vu depuis la pointe Venus. Nous distinguons la frange littorale principalement composée de cocoteraies, les forêts hygrophiles qui occupent les fonds de vallées, la végétation mésophile sur les plateaux et crêtes d'altitude et la forêt de nuages sur les plus hauts sommets. Ce paysage peut être considéré comme représentatif de la spatialisation naturelle de la végétation des îles hautes.

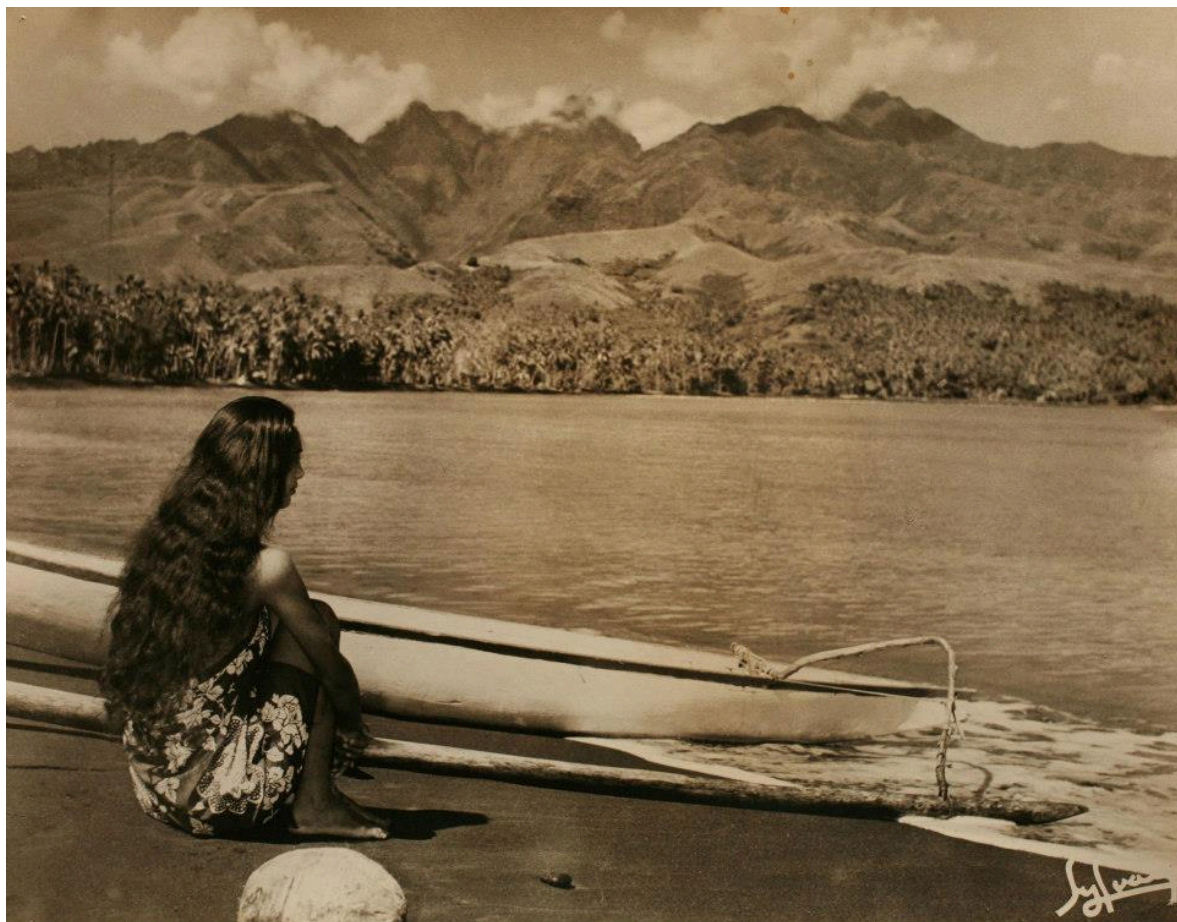


figure 4 Les hauteurs de Tahiti vues depuis la pointe Venus (Sylvain 1960)

Or, avec le temps, l'impact de l'activité humaine sur la nature s'est accru. Depuis le début des années 60, des facteurs anthropiques (urbanisation, agriculture, incendies, plantations forestières, introduction de plantes, envahissantes ou non, animaux nuisibles, etc.) ont complexifié encore davantage la spatialisation des espèces végétales.

Si l'on observe le même paysage de nos jours, on notera un net recul de la végétation naturelle au profit des zones de végétation anthropisée et de la zone urbaine qui monte de plus en plus haut. On notera ensuite que la partie naturelle du couvert végétal est graduellement envahie par une multitude d'espèces comme le pisse-pisse (*Spathodea campalunata*), le miconia (*Miconia calvescens*), les landes à fougères, le pin des caraïbes (*Pinus carribea*), le falcata (*Falcataria moluccana*), etc...

e. Espèces indigènes, exogènes et envahissantes

La Polynésie française a une longue histoire d'introductions d'espèces végétales depuis l'arrivée des premiers polynésiens, qui ont apportés des espèces principalement pour la nourriture et la pharmacopée, jusqu'à nos jours où la motivation est plus ornementale.

Le début des années 60, avec la construction de l'aéroport international de Tahiti-Faa'a en 1961 et du port autonome de Papeete en 1962, a manifestement marqué le début de l'arrivée massive d'espèces exogènes dont certaines se sont avérées envahissantes dans le contexte écologique polynésien.

Florence⁵ estime qu'il existe actuellement 890 plantes indigènes en Polynésie française (dont 570 espèces endémiques) contre 1800 espèces exogènes, ce qui illustre bien l'importance de cette problématique.

Tous ces phénomènes naturels ou anthropiques ont donc généré une complexité spatiale et une hétérogénéité de la végétation qui se traduisent par des zones monotypiques très peu étendues, et des écotones (zones de transition entre zones monotypiques) très importants. Cette mosaïque quasi-systématique d'espèces est la principale difficulté que l'on va rencontrer dans nos études sur la télédétection de la couverture végétale des îles Polynésiennes.

3. Typologie de la végétation

Pour être plus complet, détaillons ici les principaux types de couverts végétaux que l'on rencontre en Polynésie française.

La végétation des atolls est frustre, principalement constituée de cocoteraies sur un substratum corallien, avec une végétation littorale peu abondante et adaptée aux conditions difficiles rencontrées sur les atolls coralliens (peu de substrat terrigène, sable, forte concentration en sel, vents violents, érosion par la houle, ...).

En revanche, la couverture végétale des îles hautes est caractérisée par une grande diversité et une très forte hétérogénéité.

Cette végétation peut schématiquement se diviser en deux grands types (*tableau 5*) : d'une part la végétation azonale qui pousse sur certains sols ou milieux spécifiques et d'autre part la végétation zonale, dont la répartition est fonction essentiellement de facteurs abiotiques comme la température et la pluviométrie.

⁵ Florence, F., 2003 – Flore de la Polynésie française. Paris, Collection Faune et Flore tropicale. IRD, Vol. 2, 503p.

Type	Ensemble	pluviométrie	Nom
Azonale	-	-	Marécages à <i>Typha</i>
	-	-	Cocoteraies
Zonale	végétation littorale	-	-
	Végétation mésophile	1500 à 3000 mm/an	Forêts à <i>Metrosideros</i>
			Bosquet à <i>Casuarina</i>
			Bosquets à <i>Falcataria</i>
			Brousses à <i>Leucaena</i>
			Plantations à <i>Pinus</i>
			Forêts à <i>Mangifera</i>
			Landes à <i>Dichranopteris</i>
			Savanes à <i>Miscanthus</i>
	Végétation hygrophile	> 3000 mm/an	Matrice à <i>Hibiscus</i>
			Forêts à <i>Neonauclea</i>
			Faciès à <i>Aleurites</i>
			Bambouseraies
Ripisylve à <i>Inocarpus</i>			
Faciès à <i>Spathodea</i>			
Végétation orophile	3000 à 9000 mm/an	Forêt de nuages	

tableau 5 Typologie de la végétation de Polynésie française

La végétation azonale est composée des marécages à *Typha* dans les marécages salés côtiers, et des cocoteraies que l'on trouve principalement sur la plaine littorale.

La végétation zonale de son côté comporte quatre grands ensembles: La végétation littorale, la végétation mésophile, la végétation hygrophile, et la végétation orophile.

La végétation littorale est adaptée aux conditions écologiques particulières du littoral : embruns, vent, salinité, immersion. Cette végétation caractérisée par un fort taux d'endémisme subit une forte pression anthropique dans les îles habitées et, de fait, les zones de végétation littorale naturelles sont de plus en plus rares.

La végétation mésophile comprend des formations végétales semi-sèches des pentes, plateaux et crêtes de basse à moyenne altitude avec une pluviométrie comprise entre 1 500 et 3 000 mm par an. On y retrouve les forêts à *Metrosideros collina*, les bosquets à *Casuarina equisetifolia* (aito), les bosquets à *Falcataria moluccana*, les brousses à *Leucaena leucocephala* (mimosa), les plantations à *Pinus caribaea* (pins des caraïbes), les forêts anthropiques à *Mangifera zygium* (manguiers), les landes à *Dichranopteris linearis*, et les savanes à *Miscanthus floridulus*.

La végétation hygrophile comprend des formations végétales humides de fond de vallées de basse et moyenne altitude avec une pluviométrie supérieure à 3 000 mm par an. Cette végétation comprend la matrice à *Hibiscus tiliaceus*, les forêts à *Neonauclea forsteri*, les faciès à *Aleurites moluccana*, les bambouseraies à *Schizostachyum glaucifolium*, les ripisylves à *Inocarpus fagifer* (mape), les faciès à *Spathodea campanulata* (pisse-pisse).

La végétation orophile souvent qualifiée de « montagnarde » est présente sur les hauteurs des îles hautes les plus élevées avec une pluviométrie située entre 3 000 et 9 000 mm par an. Ces « forêts de nuages », fortement menacées, abritent entre 60 et 70% des plantes endémiques polynésiennes et sont le principal enjeu de préservation de la biodiversité polynésienne.

D. QUESTIONNEMENTS ET DEFIS SCIENTIFIQUES

J'ai souvent travaillé sous la forme de projets dont la durée variait de 3 mois (stages), 10 mois (stages de master) à 3 ans (thèses). Chaque projet a donné lieu à un encadrement, avec un objectif précis et des résultats attendus. Ceux-ci ont débouché sur une ou plusieurs publications. Même s'il est tentant de présenter mes résultats par projet, j'ai tenté d'organiser mon mémoire de façon thématique, l'ordre chronologique n'a donc pas été respecté. Pour aider le lecteur à trouver les détails de mes études, je cite systématiquement en fin de chapitre, ou de sous-chapitre, les articles dans lesquels sont publiés les résultats.

Toutes ces études ont eu pour but de tenter de répondre aux questions que je me suis posées au fur et à mesure de mon évolution scientifique et qui concernent chaque aspect de la chaîne de traitement de la télédétection satellitaire. Ces questions ont jalonné ma progression scientifique et je vais essayer de répondre à certaines d'entre elles dans la suite de ce document. La plupart de ces questions ont certes déjà été abordées depuis de nombreuses années, mais le milieu sur lequel je travaille est si spécifique que les réponses déjà apportées se révélaient souvent inadaptées puisque la problématique initiale était différente. Ainsi c'est l'accumulation de tous ces questionnements et de toutes les réponses que ces études ont pu apporter qui ont contribué à forger mon expérience.

J'ai organisé l'ensemble de ces questions scientifiques selon cinq axes principaux qui constitueront les cinq chapitres de mon mémoire d'habilitation à diriger les recherches :

En premier lieu, je traite la question de la définition de mon travail de recherche : quel est mon sujet de recherche ? Quels sont les moyens dont je dispose pour effectuer ma recherche ? Quels sont les facteurs contraignants et les facilités ? Sur quel milieu est-ce que je travaille ? Quelles en sont les particularités ? Quelles difficultés se posent sur ce type de milieu ?

Ce premier chapitre d'introduction présente donc le choix de mon axe de recherche qui est la télédétection environnementale en milieu insulaire tropical, et en particulier en Polynésie française.

Le contexte social y est ensuite décrit. En effet, les conditions de travail dans une université ultra-marine sont particulières, et engendrent des avantages mais aussi des inconvénients. La première d'entre elles est l'isolement scientifique qui prévaut. Sans être vécu comme un obstacle infranchissable, cette contrainte m'a malgré tout conduit à effectuer mon travail de recherche de façon très autonome et en particulier à envisager la télédétection dans toute sa transversalité. Je me suis donc intéressé à des aspects aussi variés que l'analyse des problématiques écologiques, la connaissance du milieu naturel polynésien, l'acquisition et la maîtrise des images satellites, l'expertise des algorithmes de prétraitements et de classifications, les méthodes de vérité terrain, la prise de points GPS en milieu forestier, les méthodes de fusion, la validation et la valorisation des résultats. Ainsi, mon défi résidait dans la nécessité de développer une maîtrise de l'ensemble de la chaîne de télédétection, de la définition des besoins au produit final.

Puis j'y aborde la description du contexte géographique spécifique de la Polynésie française et de ses îles. Nous l'avons vu, la végétation des îles hautes polynésiennes prend la forme d'une mosaïque complexe de couverts végétaux variés. Les différents couverts végétaux ont souvent des surfaces faibles avec des zones de transitions (les écotones) importantes. Cette mosaïque est une contrainte forte sur les techniques de télédétection traditionnelles car il ne s'agit pas de classifier des zones mono-spécifiques comme c'est souvent le cas en télédétection, mais des mélanges variés d'espèces hétérogènes. Ce point constitue le principal

enjeu scientifique auquel j'ai été confronté : comment cartographier de façon précise ce type de couverture végétale ?

Ainsi, de la même façon que le contexte social a défini mes conditions de travail et la forme qu'a prise ma recherche, le contexte géographique a été à l'origine de l'ensemble des questions et défis scientifiques que je vais développer dans les chapitres suivants.

Je me pencherai ensuite sur les images que l'on va utiliser pour détecter la végétation polynésienne. Il en existe plusieurs types : des données satellitaires optiques et radars, mais aussi des données ancillaires d'origines diverses (météorologiques, topographiques, géomorphologiques, statistiques, ...).

La problématique de l'acquisition des images satellites se pose : utiliser une image disponible en catalogue présente plus d'avantages car on en maîtrise les caractéristiques au moment de l'achat (nuages, état de surface de l'océan, ombres, ...). Cependant, les images en catalogue sont rares sur la Polynésie. Tout d'abord, les satellites ont une mémoire limitée et doivent être déchargés régulièrement pour faire de la place à de nouvelles acquisitions, or les stations de réception satellites sont rares dans le Pacifique. Ensuite, la Polynésie est isolée entre deux zones à forte potentialité commerciale que sont l'ouest du Pacifique d'un côté (avec l'Australie, la Nouvelle-Zélande et l'Asie), et les Etats-Unis et l'Amérique du sud de l'autre. Il en résulte que lors de leur passage au-dessus de la Polynésie les satellites ont soit déjà fait des acquisitions et n'ont plus la capacité suffisante pour en faire de nouvelles, soit leur capacité de mémoire est conservée pour de futures acquisitions.

Il faut alors parfois envisager de commander une image. Pour cela, il faut comprendre et déterminer les paramètres d'acquisition du satellite (taux de couverture nuageuse, angle du satellite, ...). Lors de la programmation du satellite, il faut tenter d'optimiser ces paramètres afin d'augmenter les chances d'obtenir une image satisfaisante en un temps réduit. Ce point est primordial car la couverture nuageuse régulière et persistante rend les fenêtres temporelles de prises de vues très rares et courtes.

Une fois l'acquisition effectuée et la prise de vue validée, se pose la question des prétraitements effectués par défaut par les fournisseurs d'images satellites. Ceux-ci sont-ils d'une qualité suffisante ? Ou faut-il envisager de les effectuer soi-même ? L'exemple le plus éloquent est la correction géométrique : les images satellites sont souvent livrées avec une correction géométrique basée sur le modèle numérique d'élévation mondial SRTM (NASA) au pas de 90m avec une précision altimétrique de 16m et planimétrique de 60m. Ce modèle numérique d'élévation se révèle trop imprécis pour représenter la topographie escarpée des îles hautes. Il en résulte un décalage important entre les points relevés sur le terrain et leur emplacement dans l'image, rendant la géolocalisation inexploitable. La solution consiste à fournir un modèle numérique d'élévation local avec une précision de 5m, ainsi que des points d'appuis dûment validés par le service de l'urbanisme lorsque cela est possible.

Mais la question cruciale est celle de l'intérêt de chaque type de données disponibles pour discriminer les différents couverts végétaux présents en Polynésie. Quel jeu de classes est-on capable de discriminer dans un schéma de classification mono-source avec chacune de ces données ? Leurs caractéristiques sont-elles adaptées à la résolution du problème de la complexité spatiale de la végétation ? Quelle est l'influence des spécifications techniques de ces images : définition spatiale, définition spectrales, sur la précision des cartographies finales ?

A partir de ces images, il est possible de calculer un certain nombre d'indices : ceux-ci sont essentiellement les indices de végétation et de textures tirés de la donnée optique, et les

indices polarimétriques tirés de la donnée radar. Quel est l'apport des indices usuels dans notre contexte de classification ? Il s'agit ici de bien comprendre leur intérêt pour notre chaîne de traitements et la discrimination des classes de végétations polynésiennes.

Le deuxième chapitre traite donc de l'apport des images satellitaires radars, optiques, mais aussi ancillaires.

Les apports de la donnée optique sont exposés avec deux exemples de classifications supervisées de donnée optique sur l'île de Tahiti et une étude sur l'utilisation des indices de texture pour classifier la végétation tropicale. Une seconde partie traite de l'étude de la cocoteraie des Tuamotu avec une méthode de dénombrement de cocotiers et une étude sur l'intérêt des techniques de fouille de données pour la classification globale de la base de données Ikonos.

Les apports de la donnée radar sont ensuite analysés en étudiant dans un premier temps la donnée radar en polarisations linéaires parallèle et orthogonale. Un premier exemple de classification est proposé, suivi par une étude comparative des bandes radar et un indice de végétation radar. L'apport de la polarimétrie radar pour la cartographie des milieux naturels polynésiens est ensuite étudié. Pour cela, l'apport des indices polarimétriques est analysé, puis une étude comparative des différents modes polarimétriques usuels est proposée. Une étude sur la contribution relative des indices polarimétriques en mode polarisation complète termine ce sous-chapitre.

La dernière partie de ce second chapitre est consacrée à l'apport des données ancillaires, avec une description des paramètres biophysiques permettant d'apporter une information complémentaire sur la spatialisation des espèces végétales. L'intérêt de l'utilisation des données ancillaires est étudié avec une application en télédétection indirecte sur la cartographie de la niche écologique de l'espèce envahissante *Miconia calvescens*.

La troisième partie de ma réflexion concerne les algorithmes de classification : Parmi tous les maillons de la chaîne de traitement, celui-ci me semble d'une importance capitale dès lors que l'on souhaite classifier de la donnée satellitaire ou autre.

Il existe de nombreux algorithmes, basés sur des théories mathématiques diverses, certains d'entre eux sont couramment utilisés. Il est impératif de se faire une idée précise de la pertinence de ces algorithmes dans notre contexte. Il faut par conséquent les analyser, les tester et les comparer concrètement sur nos données afin d'en appréhender précisément les avantages et les limitations éventuelles. Cela soulève un certain nombre de questions :

Pour être en mesure de comparer des algorithmes, il faut en comparer les résultats et donc utiliser une mesure de la qualité des cartes de classifications produites. Comment mesurer cette qualité ? Comment valider un résultat cartographique ?

Ensuite, parmi ces nombreux algorithmes de classification, quels sont les plus adaptés à notre problématique ? Les caractéristiques statistiques de nos images sont-elles adaptées aux hypothèses à la base des algorithmes ? Les algorithmes les plus couramment utilisés sont-ils pour autant les plus efficaces dans notre situation ? La plupart des algorithmes ont des paramètres qu'il faut régler, quelle stratégie utiliser pour optimiser ceux-ci et obtenir ainsi les meilleures précisions possibles ? Quelle est la capacité de ces algorithmes à traiter des données de grande dimension ? Le temps de calcul pour une image de grande taille est-il raisonnable ? Quel est la capacité des algorithmes à effectuer des classifications multi-sources ? Qu'en est-il si les données sont de caractéristiques hétérogènes comme dans le cas d'une classification conjointe de données optiques et radars ?

La question se pose aussi, trop souvent éludée, de la vérité terrain. Comment conduire une mission de vérité terrain de façon efficace en terrain difficile ? Quel matériel *GPS* doit-on utiliser ? Comment s'affranchir d'une topographie difficile, d'un couvert forestier dense et obtenir des points suffisamment précis pour être exploitables par rapport à la définition spatiale de nos images ? Mais surtout quelle relation existe-t-il entre les différents protocoles de vérité terrain et la précision des classifications ? Doit-on adapter le protocole de vérité terrain au type d'algorithme de classification utilisé ?

En conséquence, le troisième chapitre décrit dans un premier temps des algorithmes de classification, et pose la problématique de la mesure de la qualité des cartographies. Puis dans un second temps une comparaison multi-échelle d'algorithmes neuronaux classiques et de l'algorithme *SVM*⁶ permettra de valider l'utilisation de ce dernier sur nos problématiques.

Vient ensuite une étude plus détaillée de l'algorithme *SVM* avec un bref descriptif de celui-ci ; les techniques d'optimisation des paramètres *SVM* sont ensuite exposées.

L'utilisation des algorithmes *SVM* pour classifier de la donnée radar est validée par une comparaison avec l'algorithme de Wishart, spécialement adapté aux données radar.

Enfin, l'influence des méthodes de vérité terrain pour l'apprentissage des algorithmes neuronaux est discutée et une méthodologie adaptée aux algorithmes *SVM* est proposée.

Dès mes premières études, je suis parvenu à la conclusion que pour parvenir à cartographier la végétation polynésienne il faudrait utiliser les propriétés complémentaires de différents types de données, optique et radar par exemple. C'est l'objet de la quatrième partie de mon questionnement.

L'utilisation conjointe de plusieurs types de données entraîne une augmentation significative du nombre de bandes à traiter. De plus, ces données étant radicalement différentes par nature (imagerie passive, active), leurs propriétés statistiques sont souvent radicalement différentes. Ces deux points constituent deux limitations connues des algorithmes de classification ; comment s'en affranchir ? Nous verrons que l'algorithme *SVM* a la capacité de traiter ce genre de problèmes complexes, mais son utilisation dans un schéma de classification multi-sources est-il optimum ? Y a-t-il d'autres méthodes pour résoudre cette problématique ?

Une des techniques alternatives consiste à mettre en œuvre une fusion multi-sources. Quel schéma est le plus adapté à notre problématique ? Cette fusion sera-t-elle bénéfique pour toutes les classes ? Dans le cas contraire, que peut-on faire pour améliorer les performances ?

Le quatrième chapitre se focalise donc sur l'utilisation conjointe de données hétérogènes en analysant dans un premier temps l'intérêt de la fusion *SVM* multi-sources sur un cas concret. Puis, une version optimisée de cette dernière (fusion *SVM* sélective : *SELF*) est proposée. Ce dernier point est sans aucun doute l'apport le plus significatif développé dans ce document.

Trois applications d'utilisation conjointe de données hétérogènes sont exposées. La cartographie d'une espèce envahissante est exposée avec le cas de *Miconia calvescens*. L'utilisation et l'intérêt de ce type de technique sur la problématique des espèces rares sont explicités. Enfin, un exemple d'utilisation conjointe de données de télédétection et de *SIG*⁷ pour modéliser la santé du récif corallien illustrera ce propos.

⁶ *Séparateurs à Vaste Marge ou Support Vector Machine*

⁷ *Systèmes d'Informations Géographiques*

Pour terminer, j'examinerai mon parcours : quel bilan puis-je en faire ? Quels sont mes principaux résultats ? Comment puis-je décrire mon profil professionnel ? Qu'est ce qui fait l'originalité de ce travail ? Quels sont les perspectives qui s'ouvrent à moi pour la suite de ma carrière ? Où en est la télédétection ? Dans quelle direction va-t-elle évoluer ? Quelles sont les perspectives pour la Polynésie française ? Quels sont les secteurs où la télédétection et la géomatique vont jouer un rôle important ? Quels sont les enjeux ? Pour finir, quelles vont être les thématiques porteuses pour la recherche ? D'où viendront les perspectives de recherches intéressantes ? Quels sujets de recherche puis d'ores et déjà proposer ?

Ainsi, le cinquième et dernier chapitre de ce document présentera une vision synoptique de mon domaine de recherche. Je traiterai dans un premier temps de la synthèse de ma démarche scientifique et des principaux résultats obtenus ainsi que de mon positionnement scientifique et de l'originalité de ma démarche.

Puis, je développerai certaines perspectives que ce soit pour la télédétection en général, avec une discussion sur son état actuel et sa nécessaire démocratisation ; sur l'aspect sociétal avec une analyse des besoins et des enjeux de la télédétection ; ou bien encore sur la recherche où j'analyserai les futurs domaines porteurs au travers de l'évolution des capteurs, des techniques, des thématiques, mais aussi de sciences et de points de vues complémentaires.

Je conclurai ce mémoire avant d'exposer ma bibliographie.

II. ANALYSE DES DONNEES DE TELEDETECTION SATELLITAIRE

Pour débiter cet exposé scientifique, intéressons-nous aux images de télédétection, qui correspondent au support de l'œuvre (toile, panneau de bois, enduit en plâtre, papier, etc.) dans l'analogie artistique du préambule.

Ces images de télédétection peuvent être d'origine satellitaires mais aussi ancillaires (météorologiques, géomorphologiques ou couches *SIG*).

Les images satellites sont produites par des capteurs installés sur des satellites héliosynchrones. Ces capteurs mesurent le rayonnement électromagnétique émis par la terre, et qu'ils restituent sous forme d'images dont chaque pixel représente une valeur de ce rayonnement.

Il existe deux types de capteurs : les capteurs passifs qui mesurent un rayonnement dans une ou plusieurs longueurs d'ondes, et les capteurs actifs qui envoient une onde, à une fréquence et une polarisation définie vers la surface de la terre et mesurent ensuite la rétrodiffusion de ce signal.

Les caractéristiques des images satellites sont directement liées aux capacités de mesure des capteurs satellitaires. Elles sont principalement :

- La définition spatiale (la taille du pixel au sol) où l'on peut distinguer les images large champs (définition kilométrique), les images haute résolution (définition décamétrique) et les images très haute résolution (définition métrique et centimétrique).
- La définition spectrale (le nombre et la largeur des bandes de fréquences mesurées)
- La définition temporelle correspond au temps entre deux passages du satellite au-dessus du même endroit.

La question est de savoir quel est l'intérêt individuel de chaque type d'images (satellite, ancillaire, passif, actif, ...), l'influence de leurs caractéristiques (spatiales, spectrales, temporelles) et l'utilité des différents indices (de texture, de végétation, polarimétrique) que l'on peut en tirer, dans une optique de cartographie de la végétation.

Nous analyserons les images optiques, les images radar ainsi que les données ancillaires.

A. DONNEE OPTIQUE

La donnée optique a un intérêt certain pour la cartographie du couvert végétal. En effet, elle donne accès à l'information spectrale, à l'information spatiale par l'utilisation de la texture, ainsi qu'à une information contextuelle.

Les apports de la donnée optique haute résolution (*HR*) et très haute résolution (*THR*) sont abordés en deux temps. Dans un premier temps, la classification de données optiques à haute résolution utilisant les informations spectrales et les informations texturales sera développée. Puis dans un second temps, l'information contextuelle sera utilisée pour détecter les cocotiers des Tuamotu sur des images optiques à très haute résolution.

1. Informations spectrales et texturales

a. Classification d'images *HR* de Tahiti

Ce travail est un exemple typique de collaboration interdisciplinaire par la production d'une cartographie thématique à destination des géophysiciens travaillant sur l'érosion de l'île de Tahiti. Le cahier des charges consiste à classer une image optique de l'île de Tahiti selon un

jeu de classes simple lié à l'érosion (*i.e.* dont chaque classe puisse être liée à un coefficient d'érosion). Cette cartographie est destinée à être utilisée en entrée d'un modèle d'érosion.

Rappelons brièvement les étapes d'une classification supervisée (*cf. chapitre III.A.1*). Celle-ci commence par la mise au point d'un jeu de classes (*tableau 6*) qui décrit l'ensemble des classes que l'on souhaite identifier dans notre image.

<i>Forêt nuage</i>	<i>Eau</i>
<i>Forêt mésophile</i>	<i>Zones nues</i>
<i>Forêt humide</i>	<i>Ombre des versants</i>
<i>Plantations: Pinus et Falcata</i>	<i>Ombre des nuages</i>
<i>Landes à fougères</i>	<i>Zone urbaine</i>
<i>Cultures et pâturages</i>	<i>Nuages</i>

tableau 6 Jeu de classes Tahiti

L'étape suivante consiste à effectuer une ou plusieurs missions de vérité terrain. Celles-ci ont pour but de collecter des informations géo-référencées sur l'appartenance de certains pixels de l'image à une classe. Une base de données de vérité terrain est ainsi constituée, celle-ci est généralement scindée en deux parties, la base d'apprentissage, et la base de contrôle.

La base d'apprentissage permet à l'algorithme de classification de décider de l'appartenance d'un pixel quelconque de l'image à une classe par comparaison avec les pixels de la base d'apprentissage. La base de contrôle permet, *a posteriori*, de quantifier la qualité de la classification, présentée sous la forme d'une matrice de confusion.

Un couple d'images SPOT5 acquises en 2002 et 2009 de résolution 10 mètres en 4 couleurs (vert, rouge, proche infrarouge et infrarouge moyen) et 2,5 mètres en panchromatique a été choisi. Des masques du contour de l'île, des nuages, et de la zone urbaine (*figure 5*) ont été produits afin de restreindre la classification au milieu naturel uniquement.

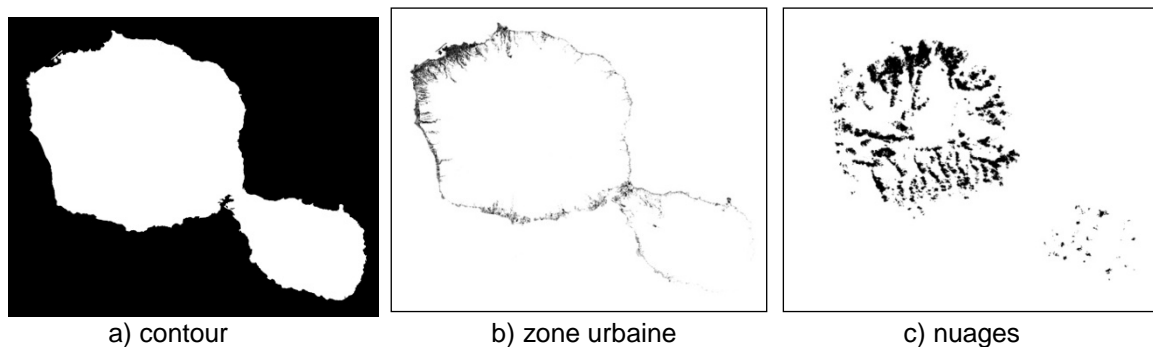


figure 5 Masques de l'île de Tahiti

Une classification supervisée de *maximum de vraisemblance* a été produite en utilisant les 4 bandes couleurs additionnées de l'élévation, la pente et l'exposition pour les images de 2002 et 2009.

Les classes *Forêt de nuages* et *Forêt mésophile* ont bénéficié d'un traitement à part puisqu'il existe une confusion forte entre ces deux classes, l'information spectrale ne permettant pas de les discriminer. Ces deux classes ont, dans un premier temps, été réunies en une seule nommée *Forêt Mésophile/Nuage* pour l'opération de classification.

Un post-traitement a permis de décider *a posteriori* si les pixels appartenaient à la classe *Forêt de nuages* ou *Forêt mésophile*. Pour cela, nous avons utilisé l'information *a priori* que nous avons sur la forêt de nuages : celle-ci est une végétation d'altitude située au-dessus de

900m pour l'ouest de Tahiti-Nui, au-dessus de 500m pour l'est de Tahiti-Nui et dès 300m sur Tahiti-Iti.

Il est important de garder la distinction entre ces deux classes car leurs caractéristiques érosives sont différentes et la forêt de nuages est un milieu de fort intérêt botanique dont la cartographie est intéressante en soi.

Une image raster composite, représentant un gradient d'altitude basé sur ces informations, a donc été créée. La classe *Forêt de nuages* est alors attribuée lorsque l'altitude d'un pixel de la classe *Forêt mésophile/nuage* est supérieure à l'altitude donnée par le gradient, et la classe *Forêt mésophile* dans le cas contraire.

Un dernier traitement a été effectué sur les résultats de classification car l'image de 2009 comportait quelques nuages. Dans ces zones, nous avons récupéré l'information de classe tirée de la classification de l'image de 2002, permettant ainsi de combler les zones sans données (figure 6).

Références : (Ye et al. *Interpraevent*'10⁸).

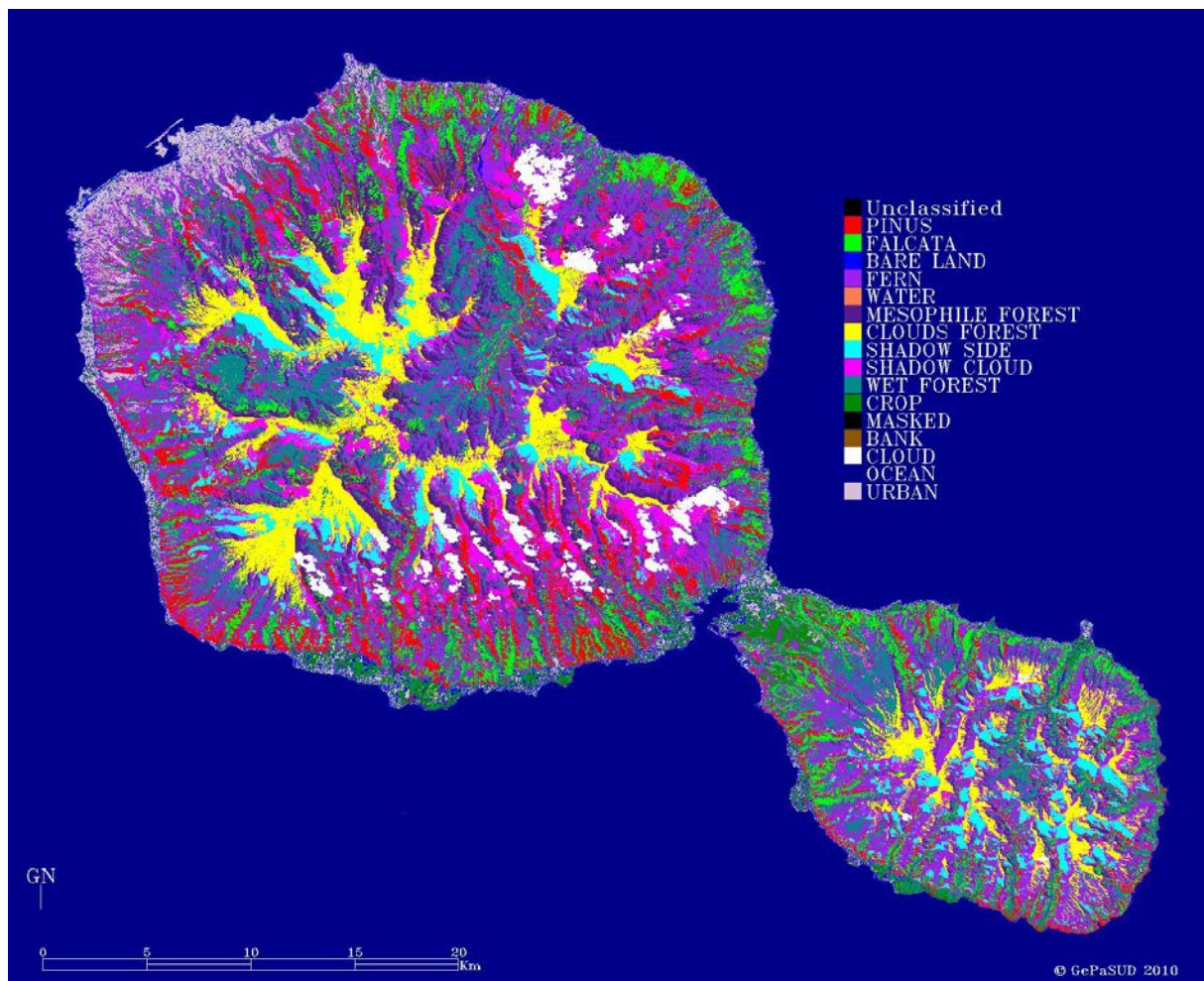


figure 6 Classification par maximum de vraisemblance de Tahiti (SPOT5 2002-2009)

⁸ F. Ye, L. Sichoix, J.-P. Barriot et P. Dumas, Modeling the Erosion of Shield Volcanoes: the Tahiti Case. International Symposium in Pacific Rim, Taipei, Taiwan, 26-30 avril 2010.p. 690-699

b. Régression de la forêt de nuages

Voici une application originale des travaux de télédétection que je conduis au sein du laboratoire GePaSud, il s'agit de la modélisation de la régression de la végétation orophile (forêt de nuages) de l'île de Tahiti.

Cet écosystème est très remarquable, on y trouve 60 à 70% des espèces endémiques des îles de la Polynésie française ; il est encore relativement préservé mais très fragile et doit être suivi et protégé. En effet, de nombreuses menaces pèsent sur lui comme l'introduction d'espèces envahissantes, l'activité anthropique mais aussi le réchauffement planétaire global.

La forêt de nuages trouve sa niche écologique au-delà d'une certaine altitude liée à la température et aux précipitations. Cette altitude va irrémédiablement augmenter avec le réchauffement global et ainsi éroder le territoire de la forêt de nuages.

Il est désormais possible de cartographier sa répartition actuelle à partir d'images optiques haute résolution comme montré au chapitre précédent (ce qui n'avait jamais été effectué par des techniques de télédétection).

Dans ce cadre, une modélisation de la régression de la forêt de nuages est envisageable en prenant en compte les paramètres du réchauffement planétaire.

La dynamique de retrait de la végétation orophile a donc été simulée par un modèle simple à partir du modèle numérique d'élévation et de la carte de la végétation orophile actuelle. La cartographie servant de base à cette étude a été obtenue par classification *SVM* des 4 bandes couleurs d'une image SPOT5 avec une définition de 10 mètres. Un modèle numérique d'élévation (*MNE*) au pas de 5 mètres obtenu par traitement de photographies aériennes stéréoscopiques a été rajouté comme couche d'information supplémentaire afin d'améliorer la qualité de la classification (ce point précis sera discuté au *chapitre III.A.2*).

L'équation suivante (*Eq. 1*) modélise l'altitude minimum de la végétation orophile sur les îles hautes polynésiennes.

$$H_t = H_i + \frac{t \cdot k}{LR} \quad \text{Eq. 1}$$

- H_t est la limite inférieure de la végétation orophile en mètres,
- t le temps en années,
- k le coefficient d'augmentation de la température ($0.0343^\circ\text{C}\cdot\text{an}^{-1}$),
- LR la diminution de la température en fonction de l'altitude ($0.0063^\circ\text{C}\cdot\text{m}^{-1}$, valeur moyenne entre 2000 et 2010)

La *figure 7* illustre la régression de la végétation orophile (en vert) de l'île de Tahiti entre 2010 et 2150 où elle aura quasiment complètement disparu. Sa superficie passe de 14 000 hectares de nos jours à 6 500 ha en 2050 et seulement 1 500 ha en 2100.

Cette application illustre bien un des intérêts de la télédétection : permettre la création de cartographies précises pouvant servir à représenter la répartition spatiale passée, présente ou future d'un milieu dans un but d'information, de prévision et de préservation.

Références : (Pouteau et al. *vertigo*2010).

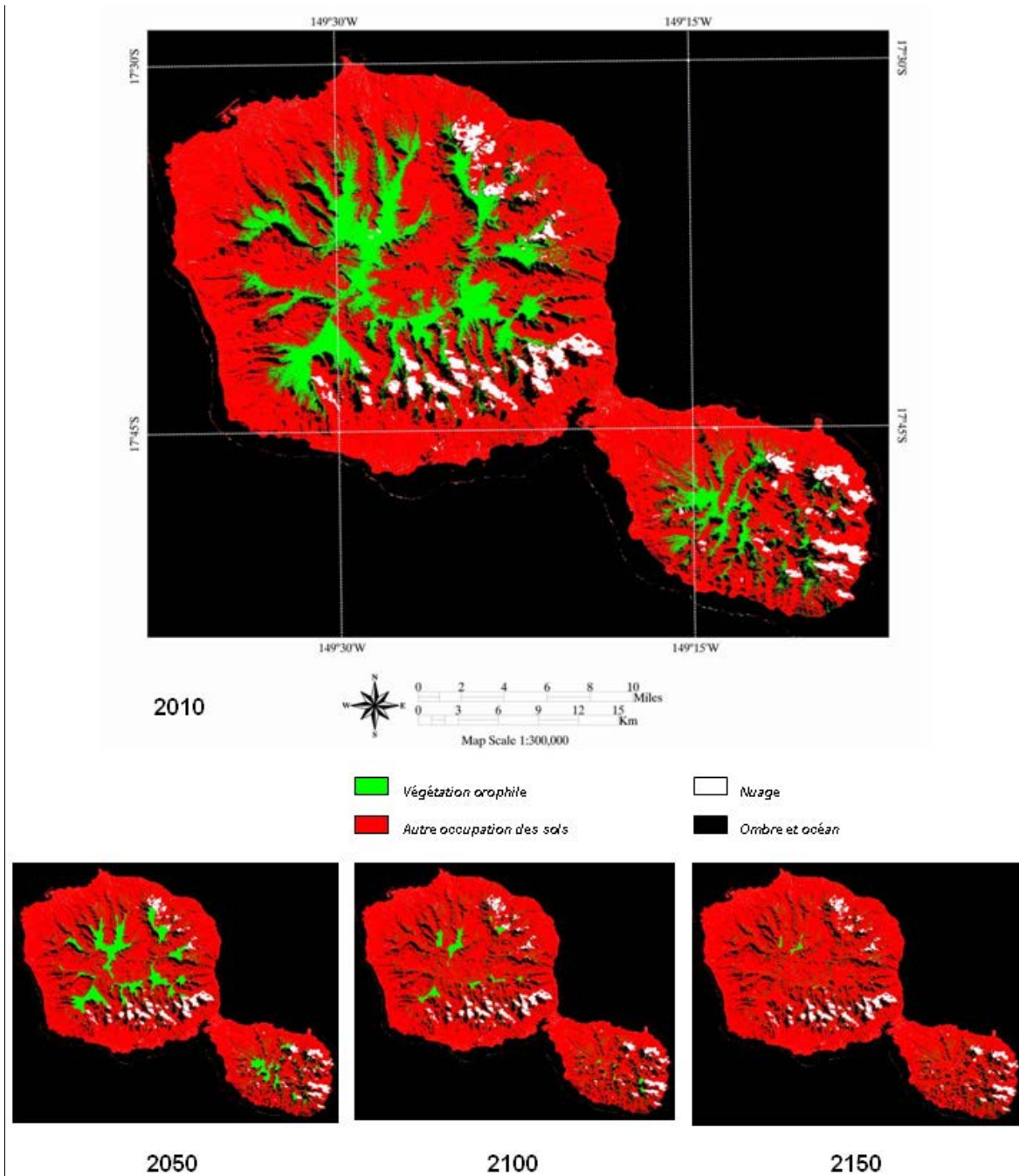


figure 7 Régression de la végétation orophile de Tahiti entre 2010 et 2150

c. Classification d'attributs de textures

La pertinence des attributs de textures pour la classification de données optiques fait l'unanimité dans le milieu scientifique puisqu'ils rajoutent une information contextuelle sur un voisinage immédiat du pixel et permettent de discriminer différents types de textures.

Lorsque l'on compare les résultats d'une classification des bandes d'une image optique et ceux d'une classification uniquement des attributs de textures tirés de cette même donnée, les performances atteintes par cette dernière sont souvent supérieures. Il y aurait donc plus d'information utile dans la texture que dans la réponse spectrale de la végétation (même si l'information spectrale est partiellement présente dans les attributs de textures classiques).

Il existe dans la littérature quantité d'attributs de textures calculés à partir des images de télédétection, les plus usités sont les attributs *GLCM* (*Gray Level Co-occurrence Matrix*)⁹.

L'utilisation de ces attributs de textures nécessite de déterminer un certain nombre de paramètres : direction, taille de la fenêtre, ... Une étude a été menée afin de déterminer une stratégie permettant de discriminer les différents couverts végétaux polynésiens par l'utilisation des attributs de textures les plus courants. Le double objectif affiché est de maximiser les performances tout en minimisant les temps de calcul.

La classification supervisée d'attributs de textures diffère un peu de la classification supervisée vue au chapitre précédent. Cette fois ci, l'algorithme classe des textures, il faut donc élaborer une base d'apprentissage de textures. Pour cela il faut constituer un jeu de classes à partir de couverts végétaux de textures différentes, et pour chaque classe prélever des échantillons géo-référencés sur le terrain. La base d'apprentissage est ensuite construite par le calcul d'attributs de textures de ces échantillons terrain.

i. Problématique de la base d'apprentissage

Calculer des attributs de textures consiste, pour chaque pixel de l'image, à calculer une certaine mesure mathématique de la texture dans un voisinage immédiat de ce pixel.

La taille de la fenêtre de calcul dépend de la texture que l'on souhaite caractériser. De façon empirique, on choisit souvent une taille égale à celle du motif de la texture la plus grossière que l'on souhaite détecter.

Le site d'étude choisi est l'île de Moorea car on y trouve l'ensemble des couverts forestiers présents en Polynésie française. Sur ce site, comme dans les autres îles hautes polynésiennes, le couvert végétal ayant la texture la plus grossière est la forêt de *Falcata* dont les couronnes peuvent atteindre 20 mètres de diamètre. Ce qui donne à peu près 32 pixels sur l'image Quickbird utilisée (4 bandes et 60cm de définition).

La constitution de la base de données d'apprentissage est soumise à deux contraintes :

Les vignettes d'apprentissages doivent être représentatives des classes que l'on essaie de discriminer, en l'occurrence, il faut chercher sur le terrain des zones homogènes (mono-spécifiques) de 20 mètres de côté.

Le nombre de vignettes par classe doit être suffisant pour éviter le problème de sous-apprentissage des algorithmes de classification, ce qui entraînerait une baisse des performances. La taille minimum de la base d'apprentissage est habituellement vérifiée de façon empirique.

Ces deux points sont des contraintes fortes dans un milieu naturel tropical fortement hétérogène. En effet, pour certaines classes, trouver sur le terrain des zones mono-spécifiques de 20 mètres de côté est problématique. En conséquence, les bases de données de vérité terrain sont souvent trop réduites.

Pour résoudre cette difficulté récurrente, étudions une stratégie consistant à sous-échantillonner les vignettes de la base d'apprentissage. Pour chaque vignette, N sous-vignettes sont extraites. Cette stratégie permet à partir d'un jeu de vignettes réduit d'augmenter considérablement le nombre de vignettes sans avoir à multiplier les zones de vérité terrain.

⁹ K. S. R. Haralick and I. Dinstein, "Textural features for image classification," IEEE Trans. Systems Man Cybernet, vol. 3, pp. 610–621, 1973.

Le calcul des attributs de textures et l'apprentissage de l'algorithme de classification se feront donc sur des vignettes de taille réduite. Cela présente un double avantage : faciliter la collecte de données terrain (moins d'échantillonnage) et permettre dans le même temps de diminuer les temps de calcul des attributs de textures et d'apprentissage des algorithmes (tous deux liés à la taille des fenêtres d'apprentissages).

Le jeu de classes est constitué de 6 couverts végétaux possédant des textures caractéristiques (tableau 7).

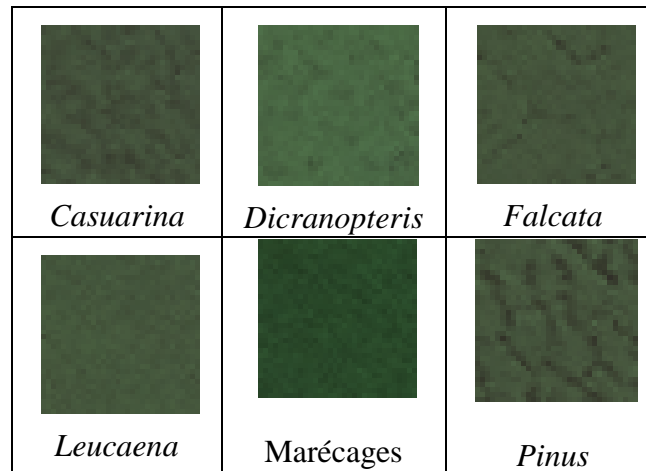


tableau 7 Jeu de classes de textures - Moorea

Une mission terrain permet de constituer deux bases de données de vignettes :

La base *B1* comprend 60 vignettes de taille 32 pixels par 32 (10 vignettes par classe).

La base *B2* comprend 3840 vignettes extraites directement de la base *B1* : pour chacune des 60 vignettes de la base *B1*, 64 vignettes de taille 25 pixels par 25 sont extraites (soit l'intégralité des vignettes possibles).

Des études préliminaires ont été effectuées pour valider l'utilisation de l'algorithme *SVM* ainsi que la stratégie d'optimisation de ses paramètres la plus pratique et la plus pertinente (cf. chapitre III.B.2).

Une comparaison des classifications produites à partir des bases d'apprentissages *B1* et *B2* est effectuée :

Dans un premier temps, les 60 vignettes de la base *B1* sont utilisées. Pour chacune d'elle 60 attributs de textures (15 attributs pour chacune des 4 bandes optiques) sont calculés. La base d'apprentissage est donc constituée de 60 images comportant 60 bandes de 32 pixels par 32.

L'apprentissage et la classification *SVM* sont réalisés sur la totalité de la base *B1*.

Le taux de bonne classification *GRP*¹⁰ correspond au nombre de vignettes bien classifiées divisé par le nombre total de vignettes. Celui-ci nous donne une mesure de la qualité de la classification. Le *GRP* obtenu est de 93% dans ce cas de figure.

Dans un second temps, la base *B2* comportant 3840 vignettes de 15 pixels par 15 est scindée en deux : la moitié (1920 vignettes) est utilisée pour l'apprentissage et l'autre moitié est classifiée. Dans ce cas, le taux *GRP* est de 100%.

Cette expérience montre que cette technique de sous-échantillonnage des vignettes de vérité terrain permet d'une part d'utiliser des vignettes d'apprentissage de dimensions plus petites

¹⁰ *Good recognition percentage*

que la texture à classifier entraînant de facto un gain des performances de la classification, et d'autre part de diminuer le nombre de zones d'apprentissage à échantillonner.

Cependant, l'utilisation de la base *B2* augmente considérablement le nombre d'images pour le calcul des attributs de textures et l'apprentissage, ce qui entraîne une forte augmentation du temps de calcul.

L'étape suivante consiste alors à vérifier s'il est possible de diminuer la quantité de vignettes d'apprentissage tout en conservant des performances équivalentes.

L'expérience consiste donc à diminuer le nombre de vignettes de la base *B2* et à relever le *GRP* en fonction du nombre de vignettes utilisées. Les résultats de la *figure 8* montrent qu'un bon compromis est obtenu lorsque l'on utilise 30 vignettes par classe (180 vignettes au lieu des 1920 vignettes initiales) avec lequel on atteint un *GRP* de 97%.

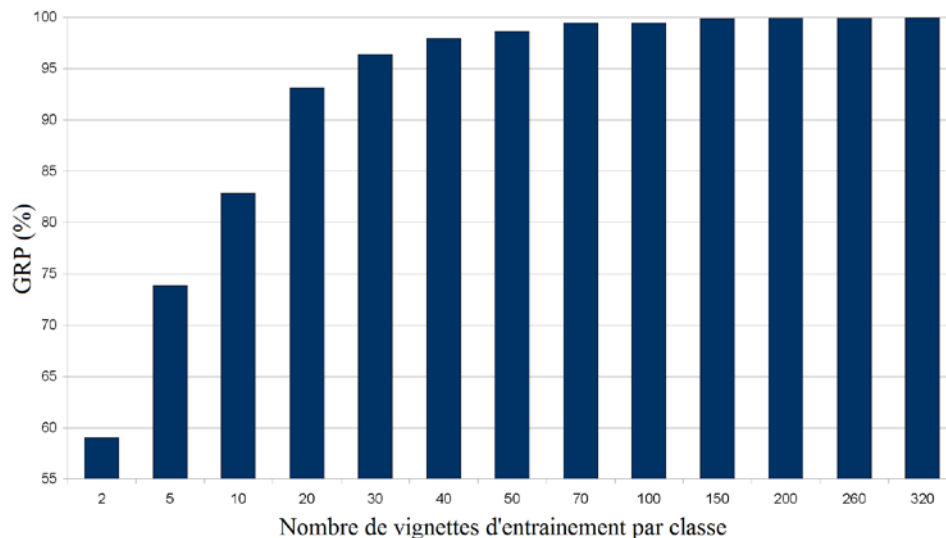


figure 8 Influence du nombre de vignettes d'apprentissage

ii. Eliminer les redondances parmi les attributs de textures

L'autre facteur que l'on souhaite minimiser est la quantité d'attributs de textures. En effet, certains attributs de textures parmi les 60 précédemment calculés peuvent être redondants ou non pertinents. Dans ce cas, se limiter aux attributs de textures les plus pertinents devrait permettre d'améliorer les performances tout en réduisant les temps de calculs.

Pour cela, deux classifications sont opérées avec un apprentissage sur 60 attributs (15 attributs pour chacune des 4 bandes optiques) et sur 30 attributs (uniquement sur les bandes verte et infrarouge).

Dans un second temps, une analyse en composante principale (*ACP*) a été opérée sur l'ensemble des 60 attributs de textures en vue d'en réduire le nombre. Des classifications avec apprentissage sur les 10, 8, 6, et 4 premiers attributs ont été conduites. Les résultats de ces dernières sont comparés aux résultats des deux premières classifications (*figure 9*).

La conclusion de cette expérience est que l'utilisation de l'ensemble des 60 attributs obtient un *GRP* de 10% inférieur à un apprentissage sur les 8 ou 10 premiers attributs de textures, ce qui confirme l'hypothèse que certains attributs sont dispensables. Cette technique permet de s'en affranchir et donc d'atteindre de meilleures performances.

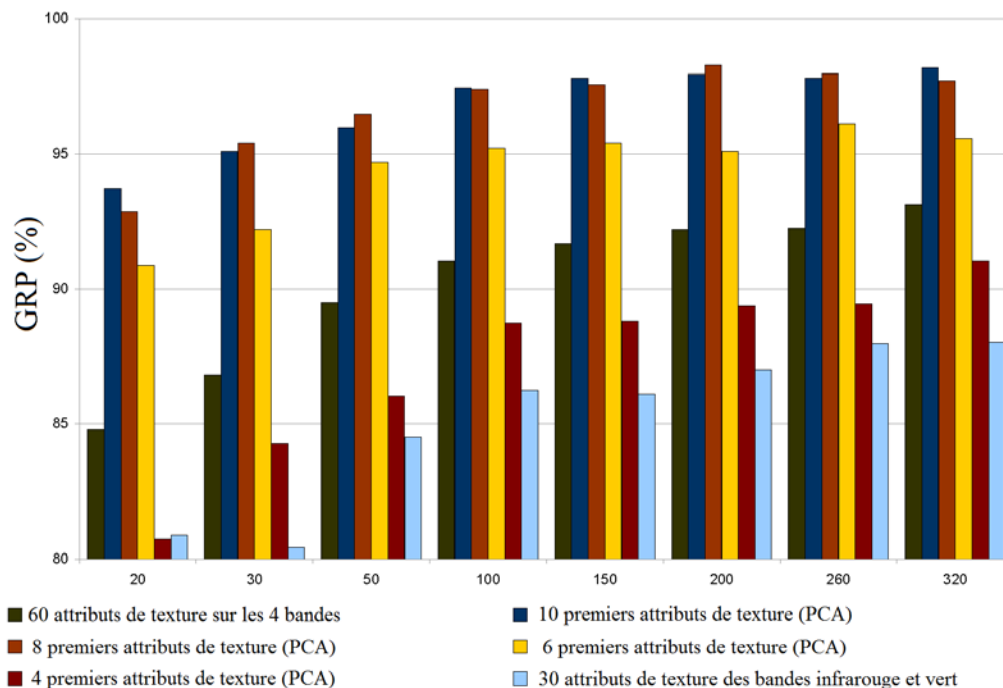


figure 9 Influence du nombre d'attributs de textures

Voici succinctement les conclusions globales de cette étude :

- Lorsque les missions de vérité terrain sont difficiles et le milieu hétérogène, il est possible de compenser le faible nombre de zones de vérité terrain grâce à la technique de sous-échantillonnage présentée.
- La définition d'une trentaine de vignettes par classe permet d'obtenir un bon compromis entre le temps d'apprentissage et les performances atteintes.
- Grâce à un choix judicieux des attributs de textures, nous pouvons en réduire le nombre de 60 à une dizaine tout en améliorant les performances de la classification.

Références : (Chabrier et al. aprs2012).

2. Étude de la cocoteraie des Tuamotu

L'un des milieux naturels polynésiens les plus typiques est la cocoteraie de l'archipel des Tuamotu. Cette cocoteraie occupe la quasi-totalité de la surface immergée des 78 atolls coralliens, elle est d'un intérêt primordial pour l'économie de la Polynésie française. Cependant, force est de constater qu'elle est aujourd'hui sous exploitée, la culture du coprah ayant été peu à peu délaissée au profit d'emplois moins pénibles et plus rentables, tout particulièrement au moment de l'arrivée du CEP (Centre d'Essais du Pacifique).

Le service du développement rural (SDR) du territoire de la Polynésie française a entre autres mission de relancer la filière coprah, ainsi que de développer l'exploitation du bois des cocotiers séniles. A cette fin, il est demandeur de données qualitatives et quantitatives concernant la cocoteraie. Des biologistes conduisent régulièrement des missions de terrain sur certains atolls considérés comme prioritaires afin d'estimer les surfaces et l'état de la cocoteraie pour mettre en place un plan de relance de l'exploitation du coprah.

Toutefois, compte tenu du nombre important d'atolls et de leur accessibilité limitée, obtenir une idée précise et exhaustive des surfaces de cocoteraies semble une tâche ardue. L'outil télédétection se révèle alors particulièrement adapté pour répondre à ce besoin, d'autant plus

qu'une base de données d'images Ikonos quasi complète des atolls des Tuamotu est disponible au service de l'urbanisme du territoire de la Polynésie française.

a. Dénombrement des cocotiers

Des études ont montré la faisabilité de la détection individuelle des couronnes d'arbres à l'aide de données optiques à très haute résolution, sur des champs d'oliviers notamment. Ces études nous permettent d'envisager le même type de traitements pour détecter les cocotiers.

L'objectif *in fine* est de repérer tous les cocotiers sur une image, de les compter et d'analyser leur distribution spatiale afin de fournir des données précises.

Le site d'étude est l'atoll de Tikehau qui fait partie de la liste des atolls prioritaires pour le service du développement rural. Cet atoll comporte un *motu* (îlot corallien) principal, où se situe le village, et des *motu* périphériques, où les cocoteraies sont exploitées par les habitants.

i. Segmentation des cocotiers

La première étape consiste à créer un masque pour ne conserver que la cocoteraie, celui-ci est réalisé en deux temps :

- Tout d'abord, un masque ne gardant que la végétation de façon générale est créé. La donnée Ikonos à notre disposition ne comportant que trois couleurs (rouge, vert, bleu), une classification par *maximum de vraisemblance* permet de s'affranchir de l'absence de la bande infrarouge. Un jeu de classes simple permet de discriminer les pixels considérés comme appartenant à la classe végétation. Une opération morphologique de fermeture-ouverture permet ensuite d'homogénéiser le résultat.
- Parallèlement, une analyse de la texture de l'image nous permet de différencier la végétation haute (principalement constituée de cocotiers) de la végétation moyenne et basse. L'utilisation conjointe de ces deux masques permet de ne conserver que la cocoteraie. Cette étape est essentielle pour la suite afin d'éviter les fausses détections.

La procédure consiste ensuite à repérer chaque couronne d'arbre dans l'image.

Il existe plusieurs algorithmes de segmentation des couronnes d'arbres. Citons la recherche de maxima locaux du Laplacien, la recherche de motifs, le suivi de vallée, l'algorithme de délimitation individuelle des arbres *TIDA*, l'algorithme de ligne de partage des eaux et la technique dite des processus ponctuels marqués (*Teina thesis2009*).

La technique choisie est celle dite de ligne de partage des eaux (*watershed*). Le principe repose sur le fait que les arbres éclairés par le soleil ont un sommet plus brillant que le pourtour qui est généralement plus sombre, il s'agit donc de segmenter l'image en suivant ces vallées plus sombres entre les arbres.

L'image en niveaux de gris est considérée comme une surface, celle-ci comporte des vallées et des sommets. L'algorithme « inonde » les vallées à partir de leur point le plus bas et lorsque l'eau atteint une crête donnant sur autre vallée, l'emplacement de la crête est préservé et constitue la frontière entre deux zones à segmenter.

En pratique, après avoir inversé l'image, l'algorithme est initialisé en cherchant les minimas locaux de l'image (marqueurs). Les marqueurs correspondent aux sommets des cocotiers (*figure 10*).

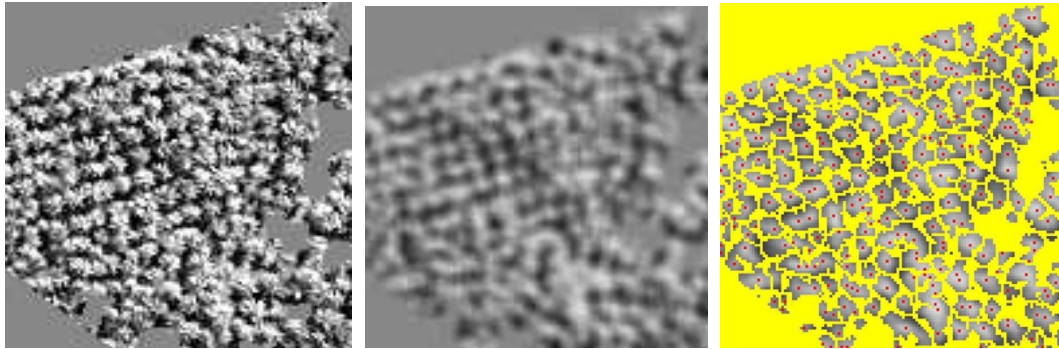


figure 10 Technique de localisation des marqueurs

Le résultat de cette étape est une segmentation de toutes les couronnes visibles des arbres de la cocoteraie (figure 11). Une étape de post-classification élimine les zones dont la surface est trop importante ou trop faible pour être un cocotier. La surface d'un cocotier adulte se situe dans une fourchette comprise entre 12 et 120 m².

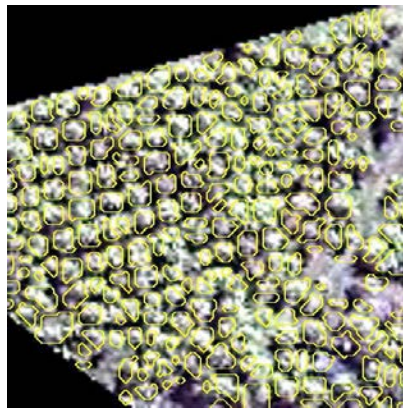


figure 11 Segmentation des couronnes des cocotiers

Cette procédure nous permet d'obtenir un ensemble de zones segmentées dont les centres géométriques forment un nuage de points représentant chacun le sommet d'un cocotier.

ii. Validation de la méthode

Pour valider les résultats précédents, un protocole de validation a été mis au point en plusieurs étapes :

Tout d'abord un repérage manuel des cocotiers sur l'image : Un certain nombre de personnes pointent chaque cocotier discerné sur l'image. Puis nous avons comparé le nombre de cocotiers trouvés par notre méthode de segmentation avec celui trouvé par les opérateurs humains.

Le résultat de ce test s'est avéré concluant puisque la méthode par segmentation et la méthode par photo-interprétation humaine ne diffèrent jamais de plus de 10%. Cela permet d'affirmer que l'on détecte les cocotiers visibles sur l'image avec une marge d'erreur faible.

Ensuite, une mission de vérité terrain a été effectuée sur l'atoll de Tikehau afin de valider *in situ* les résultats. Deux protocoles ont permis la comparaison entre les résultats obtenus en laboratoire et la réalité sur le terrain.

Le premier protocole est un dénombrement exhaustif des cocotiers présents sur un petit *motu*, (*motu* 16) afin de comparer avec le nombre de cocotiers obtenu par notre méthode.

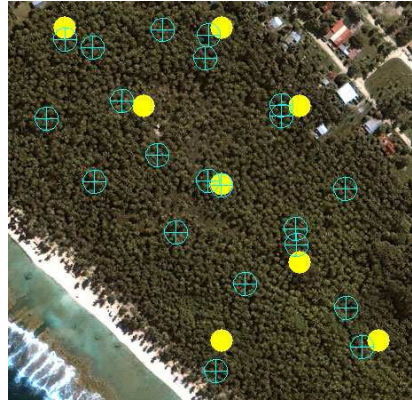


figure 12 Échantillonnage terrain prévu et effectué

Le second protocole consiste à échantillonner la cocoteraie de façon systématique selon un maillage régulier de placette circulaire d'un rayon de 12 mètres, la *figure 12* représente les échantillons prévus en jaune, et ceux effectués en croix. Dans chaque placette nous comptons le nombre de cocotiers afin de comparer la densité de cocotiers réelle avec celle trouvée sur le nuage de points obtenu par segmentation.

Sur le terrain, nous avons réalisé que le nombre de cocotiers comptés sur le terrain était beaucoup plus important que celui trouvé par notre algorithme.

Cela s'explique par le fait que notre algorithme ne compte que les cocotiers visibles sur l'image. Ce point est corroboré par la validation précédente.

Une stratification verticale de la cocoteraie considérant 3 strates a été adoptée (*figure 13*) :

- S1 : les cocotiers en canopée (visibles par le satellite imageur),
- S2 : les cocotiers partiellement recouverts par des cocotiers plus hauts (partiellement visibles par agglomération avec la strate S1),
- S3 : les cocotiers totalement recouverts par d'autres cocotiers (invisibles sur notre image).

Le protocole terrain a été adapté en recensant le nombre de cocotiers présents dans chaque strate pour chaque station d'échantillonnage.

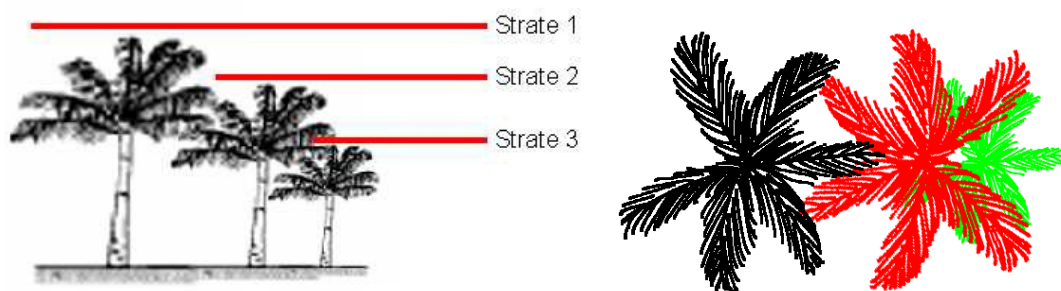


figure 13 Strates échantillonnées des cocotiers

Les résultats montrent que la méthodologie mise au point est adaptée puisque tous les cocotiers de la strate S1 du *motu 16* (échantillonnage exhaustif) sont détectés avec une erreur de 4% seulement. Si l'on considère l'ensemble des zones échantillonnées de l'atoll de Tikehau, l'erreur de détection de la strate S1 varie entre 2 et 5%.

Cependant, les cocotiers des strates S2 et S3, bien que non-visibles doivent être pris en considération. Un taux de détection des cocotiers a été calculé comme le rapport du nombre de cocotiers détectés par le nombre de cocotiers effectivement présents sur le terrain.

Ce taux de détection est proche de 60% sur l'ensemble de Tikehau. Il est fortement influencé par le type de cocoteraie : celui d'une cocoteraie plantée de façon régulière, espacée, entretenue, sur un sol sablonneux est très important alors que celui d'une cocoteraie sauvage ou abandonnée est plus faible.

Pour estimer de façon plus précise le nombre effectif de cocotiers, une typologie des cocoteraies est proposée. Un taux de détection spécifique à chaque type de cocoteraie est attribué par des mesures sur le terrain.

Les trois types de cocoteraie que l'on peut observer sur l'atoll de Tikehau (*figure 14*) sont :

- En jaune : la cocoteraie sauvage ou naturelle,
- En cyan : la cocoteraie artificielle (plantée) à basse densité (entretenu),
- En violet : la cocoteraie artificielle à haute densité (non entretenue)



figure 14 3 types de cocoteraies

La suite logique de ce travail consiste à segmenter la cocoteraie selon ces trois types. Deux méthodes ont été explorées : la classification d'attributs de textures et l'analyse spatiale du nuage de points des couronnes d'arbres.

iii. Segmentation de la cocoteraie par classification

Une première approche consiste à classifier la cocoteraie. Des essais ont montré que l'utilisation des bandes optiques seules ne permettait pas une discrimination suffisante, mais que l'ajout de l'information de texture permet d'obtenir des résultats probants.

Une étude sur la taille optimale des fenêtres de calcul des attributs de textures a permis de fixer la taille à 19 par 19 pixels.

Une comparaison (Teina *et al.* spie2009) des principaux algorithmes de classification a été effectuée, montrant que les deux algorithmes obtenant les meilleurs résultats sont les algorithmes de *maximum de vraisemblance*, et celui de *séparateur à vaste marge (SVM)*, loin devant les algorithmes de classification par *distance de Mahalanobis*, par *parallélépipède*, et par *distance minimale*.

L'algorithme de *maximum de vraisemblance* montrant une saturation de ses performances pour des tailles de fenêtre de texture trop grandes, l'algorithme *SVM* est choisi (*figure 15*).

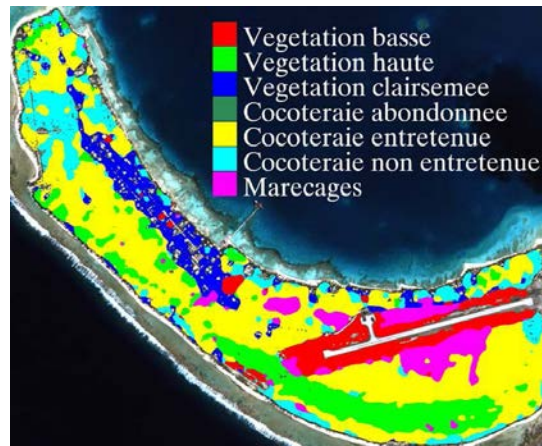


figure 15 Classification SVM de la cocoteraie du motu principal de Tikehau

iv. Segmentation de la cocoteraie par analyse spatiale

La seconde approche consiste à analyser le nuage de points précédent par une analyse spatiale morpho-mathématique.

Les cocoteraies naturelles (sauvages) sont modélisées par un processus ponctuel de Poisson. Si le motif du nuage de points ne correspond pas à celui-ci, la cocoteraie est considérée comme artificielle. Dans le cas contraire, elle est considérée comme naturelle.

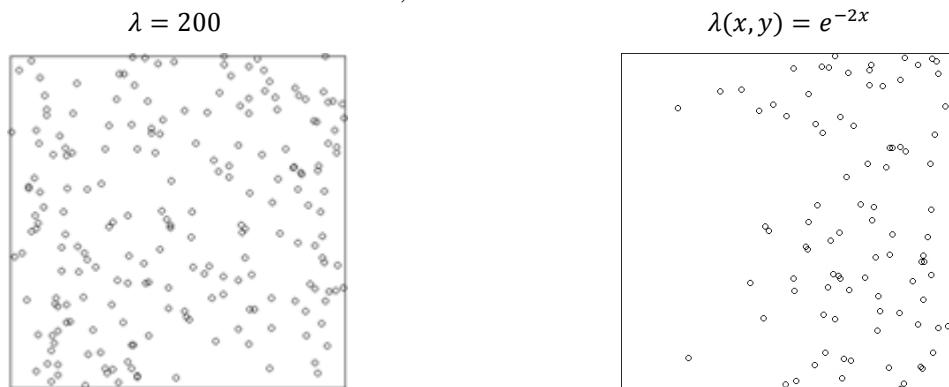


figure 16 Processus ponctuels de Poisson homogène et non-homogène

Nous cherchons donc à segmenter le nuage de points par rapport à un critère d'homogénéité basé sur la densité des cocotiers, densité directement liée aux paramètres de la distribution de Poisson. La figure 16 représente une simulation d'un processus de Poisson homogène (à gauche) et non-homogène (à droite), dans ce cas, la densité devra être estimée localement.

Plusieurs méthodes ont été testées pour estimer la densité, l'analyse par quadrants, la méthode des noyaux et l'analyse du second ordre. Après tests et comparaisons, la méthode des noyaux a été utilisée pour estimer la carte des intensités (figure 17).

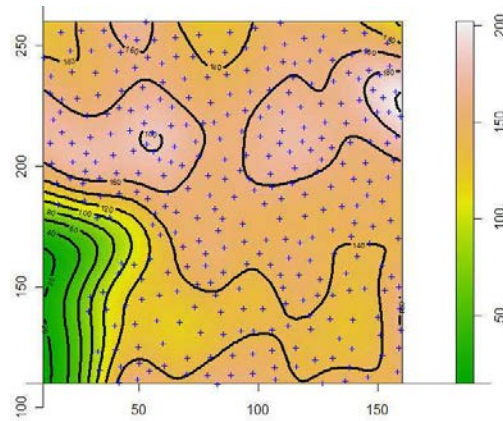


figure 17 Estimation de λ pour une sous-région du nuage de points réel.

Cette carte de densité est segmentée en 3 classes grâce à un seuillage de l'histogramme des densités. Les seuils tiennent compte de la valeur théorique de densité des cocotiers.

En effet, pour l'espèce présente sur l'atoll de Tikehau, les cocotiers *Grands Locaux*, l'espacement optimum recommandé par le Service du Développement Rural est de 8m50, ce qui nous donne une densité de 140 à 150 arbres par hectare.

Le seuillage de la carte de densité (figure 18 à gauche) permet la segmentation finale de la cocoteraie (figure 18 à droite).

Par rapport à la méthode de classification précédente, cette méthode a pour principal avantage un gain de temps non négligeable, mais, elle nécessite d'avoir accès au nuage de points issu de l'étape de segmentation individuelle des cocotiers.

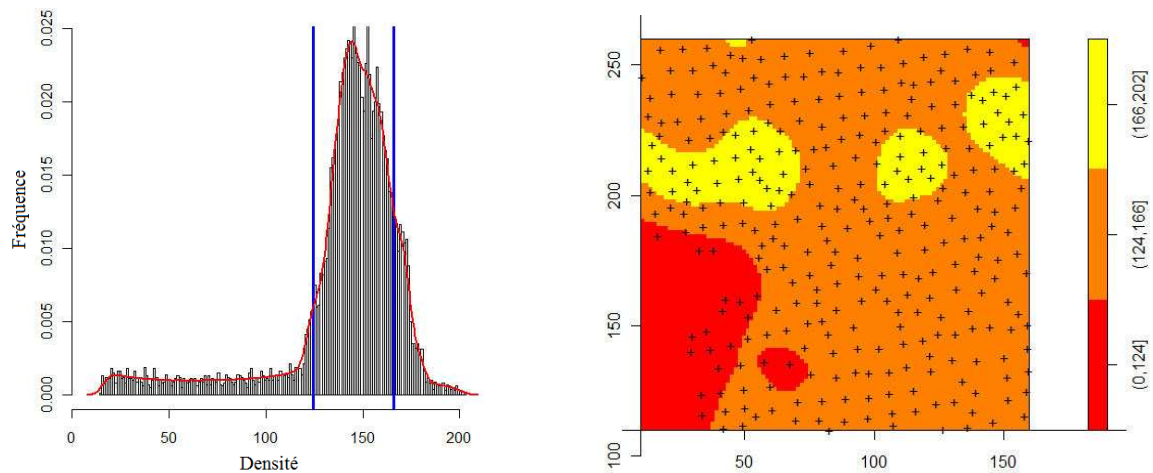


figure 18 Histogramme des densités λ et segmentation de la cocoteraie

v. Résultats finaux

Le rapport entre le nombre de cocotiers détectés et ceux présents sur le terrain (données issues de notre campagne terrain et de celles du *SDR*¹¹ permet d'assigner un taux de détection moyen à chacune des 8 classes de la classification *SVM* de la figure 15.

¹¹ Frédéric Jacq -2006

	Végétation basse	Végétation haute	Végétation épaisse	Cocoteraie abandonnée	Cocoteraie entretenue	Cocoteraie non entretenue	Marécages	Pas de végétation	
Taux de détection moyen	92,70%	64,44%	0%	66,76%	72,75%	61,11%	57,04%	0,00%	Total
DéTECTÉS	13 771	21 735	650	87 824	23 439	1 334	973	904	150 630
EXTRAPOLÉS	14 855	33 728	650	131 546	32 220	2 183	1 706	904	217 792

tableau 8 Extrapolation finale du nombre de cocotiers sur l'atoll de Tikehau

Le nombre total de cocotiers présents sur l'atoll de Tikehau est ainsi estimé par extrapolation (tableau 8). Il est de 150 630 cocotiers détectés, et 217 792 cocotiers estimés.

Références : (Teina et al. majecstic2007; Teina et al. spie2008; Teina et al. igarss2008; Teina et al. spie2009).

b. Fouille de données

Aucune des deux méthodes de segmentation du chapitre précédent n'est adaptée pour traiter l'intégralité des images Ikonos de l'ensemble des atolls des Tuamotu (59 principaux atolls, 306 scènes Ikonos, soit 150Gb de donnée). L'intervention humaine serait trop importante pour les prétraitements et les apprentissages, sans compter la longueur des temps de calculs.

Nous souhaitons orienter ici nos recherches vers une méthode semi-automatique pour classifier la cocoteraie sur la base de données complète d'images Ikonos des Tuamotu. Cette classification ayant pour seul but de connaître la superficie globale de la cocoteraie, une seule classe de cocoteraie est considérée.

Une collaboration avec l'équipe pôle pluridisciplinaire de la matière et de l'environnement (PPME - EA n°3325) de l'université de la Nouvelle-Calédonie a été entamée afin de tester les techniques de fouille de données sur cette base de données.

0	Non-Classifié	6	Corail Jaune
1	Cocoteraies	7	Corail Gris
2	Végétations Autres	8	Corail Sombre
3	Eaux Profondes	9	Sable Blanc
4	Eaux Peu Profondes	10	Vagues
5	Corail Vert	11	Nuages

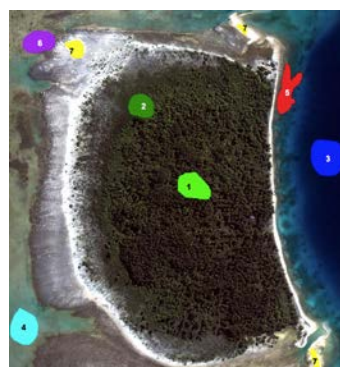


tableau 9 Jeu de classes - Tuamotu

figure 19 Exemples des différentes classes

La première piste envisagée pour réduire au maximum l'intervention humaine est de réduire les prétraitements. Ainsi, contrairement aux classifications précédentes, la zone d'étude n'est pas réduite par masquage, ce qui conduit à intégrer dans le jeu de classes (tableau 9) des classes représentant les plages, le lagon et l'océan. Un exemple de ces classes est représenté sur la figure 19.

La seconde piste pour réduire les temps de traitement consiste à créer des zones d'apprentissages pour des groupes d'images ayant des paramètres de prise de vue similaires, dans lesquelles les classes possèdent des caractéristiques équivalentes.

Les paramètres influant sur les caractéristiques des images se trouvent pour la plupart dans les fichiers de métadonnées. Ainsi les métadonnées permettent d'accéder à la date, l'heure d'acquisition, la position du satellite, du soleil, etc.

Ces paramètres n'influent pas tous sur les caractéristiques des images, seuls seront considérés les plus pertinents. Par exemple ceux associés aux conditions d'illumination comme l'azimut et l'élévation du soleil, mais aussi la position du satellite (azimut et élévation) qui, associés à la position du soleil, expliquent la brillance de certains objets comme l'océan, le lagon ou les palmes de cocotiers.

D'autres données directement issues des images sont analysées, comme l'histogramme qui permet de regrouper des images ayant une dynamique et un contraste équivalent.

Une méthode de clustering *X-Mean* a été appliquée sur les sept métadonnées choisies. Cette méthode est équivalente à la méthode de *K-Mean* mais le nombre de groupes (Clusters) est trouvé automatiquement alors qu'il est fixé par l'utilisateur dans la méthode *K-Mean*.

Pour chaque groupe d'images déterminé, une analyse de la valeur centroïde des métadonnées permet de trouver l'image qui possède les caractéristiques les plus proches du centroïde du groupe. Cette image est considérée comme représentative de son groupe. Elle est utilisée pour définir les zones d'apprentissage (*ROI*), ces zones sont utilisées pour la classification de l'ensemble des images du groupe.

Dans certains cas, des classes sont absentes de l'image de référence, une autre image proche du centroïde est alors choisie (la deuxième, troisième, etc.). Le but étant d'utiliser une image possédant l'intégralité des classes afin d'éviter les erreurs de classifications.

L'algorithme de classification utilisé est un algorithme d'*arbre de décision (Decision Tree)*, l'implémentation *J48* de l'algorithme *C4.5* a été choisie. Une décomposition binaire des composantes *RVB* a été utilisée pour améliorer les performances de l'algorithme. La *figure 20* illustre la méthode proposée.

Certes, les résultats de classification obtenus sont moins bons que pour une classification supervisée classique, mais ils demeurent prometteurs et les temps de calculs sont plus adaptés à la base de données considérée.

Il reste néanmoins à améliorer la qualité des classifications en analysant plus précisément l'influence de chaque paramètre extrait des métadonnées associées, tester de nouvelles méthodes de clustering, etc...

Cette piste de recherche n'en est qu'à ses débuts et devra être développée dans les années à venir afin d'améliorer la précision des classifications.

Références : (Desmier et al. icdim2011).

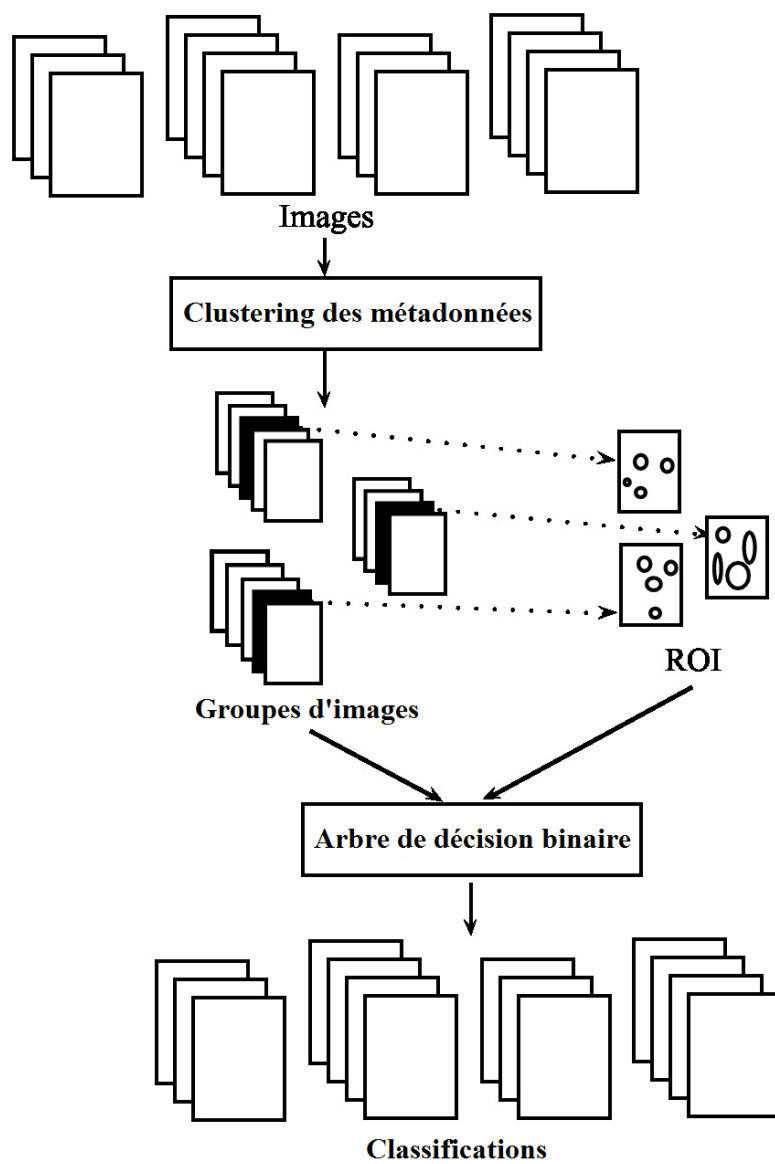


figure 20

Méthode de classification par fouille de données

B. DONNEE RADAR

Comme évoqué précédemment, en milieu tropical insulaire, la couverture nuageuse quasi systématique et persistante rend très difficile l'obtention d'images optiques exploitables. Les capteurs radar étant actifs, ils ont la capacité de prendre des images de nuit comme de jour et sont peu sensibles à la présence de nuages.

L'apport des données radar est analysé en étudiant dans un premier temps la donnée radar en polarisation linéaire simple (polarisations parallèles et orthogonales), puis dans un second temps la donnée radar polarimétrique.

1. Donnée radar en polarisation linéaire simple

a. Classification mono-source

La campagne PACRIM II du bureau JPL de la NASA a eu lieu en 2000 avec pour objectif d'effectuer une acquisition d'images radar sur les principales îles polynésiennes à l'aide d'un capteur radar aéroporté AirSAR : Les premières données ont été délivrées en 2001 et concernaient l'île de Moorea. Ces données m'ont permis d'effectuer une étude préliminaire sur l'intérêt de la classification supervisée d'images radar pour la cartographie de l'occupation des sols et du couvert végétal en particulier.

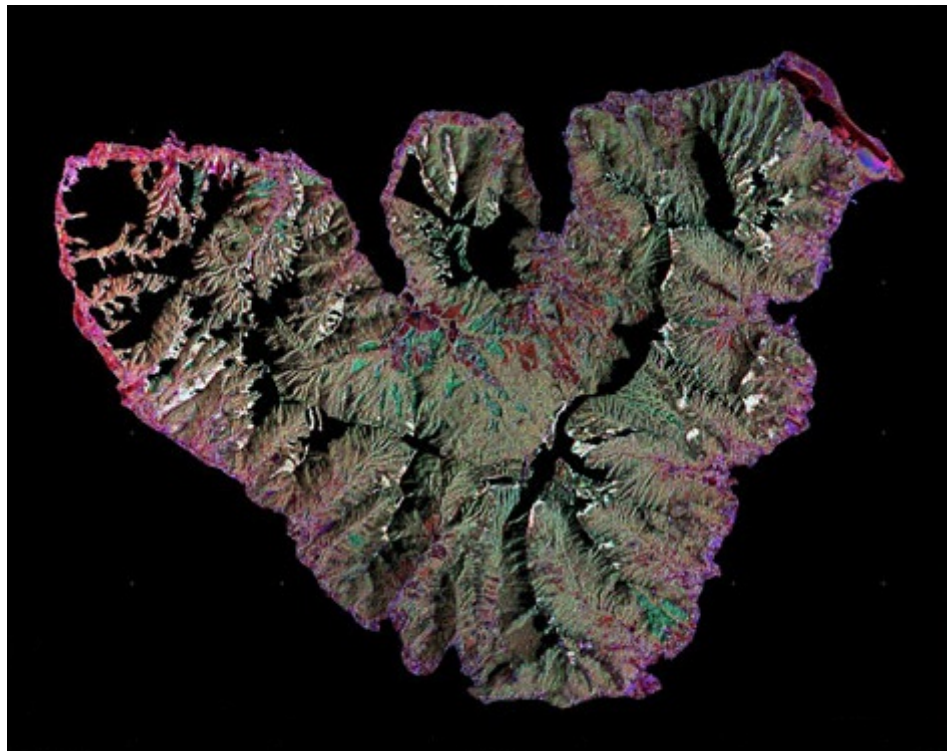


figure 21 Composite AirSAR (C_{VV} , L_{HV} , L_{HH}) - Moorea

L'observation de la donnée AirSAR de l'île de Moorea (figure 21) confirme bien l'insensibilité aux nuages, mais cette image présente des zones sans donnée. Cela est dû au caractère aéroporté du capteur, qui, orienté sur le côté de l'avion n'a pas de visibilité sur la face cachée des reliefs accentués.

Toutes les prises de vues ont par conséquent été doublées avec deux passes symétriques afin de permettre éventuellement de combler les trous par mosaïquage. Ce n'est pas le cas de Moorea, pour laquelle la fauchée du satellite étant plus petite que sa taille, les deux passes symétriques ont servi à couvrir l'ensemble de l'île avec très peu de recouvrements.

La donnée JPL-AirSAR fournit des images en bande C-TopSAR ($C_{VV} + MNE^{12}$) et en bande L-PolSAR et P-PolSAR (quadruple polarisation VV, HV, HH, VH). La définition spatiale est de 5m.

Notons que lors du survol de l'île de Moorea, la bande P_n n'a pu être acquise. Les bandes radar analysées sont la bande C_{VV} ainsi que les polarisations simples de la bande L : L_{VV} , L_{HH} , et L_{HV} .

Le site d'étude choisi est la caldera centrale de Moorea (domaine d'Opunohu), car sa topographie relativement homogène permet de s'affranchir dans une moindre mesure des problèmes d'inversion de pentes inhérents à la donnée radar.

Le domaine d'Opunohu (figure 22) est une zone d'intérêt économique de première importance en Polynésie. Celle-ci présente une palette de types d'occupation des sols hautement représentative des îles hautes polynésiennes : différents types de forêts, cultures, pâtures, étendues d'eau, zone urbanisée, etc.

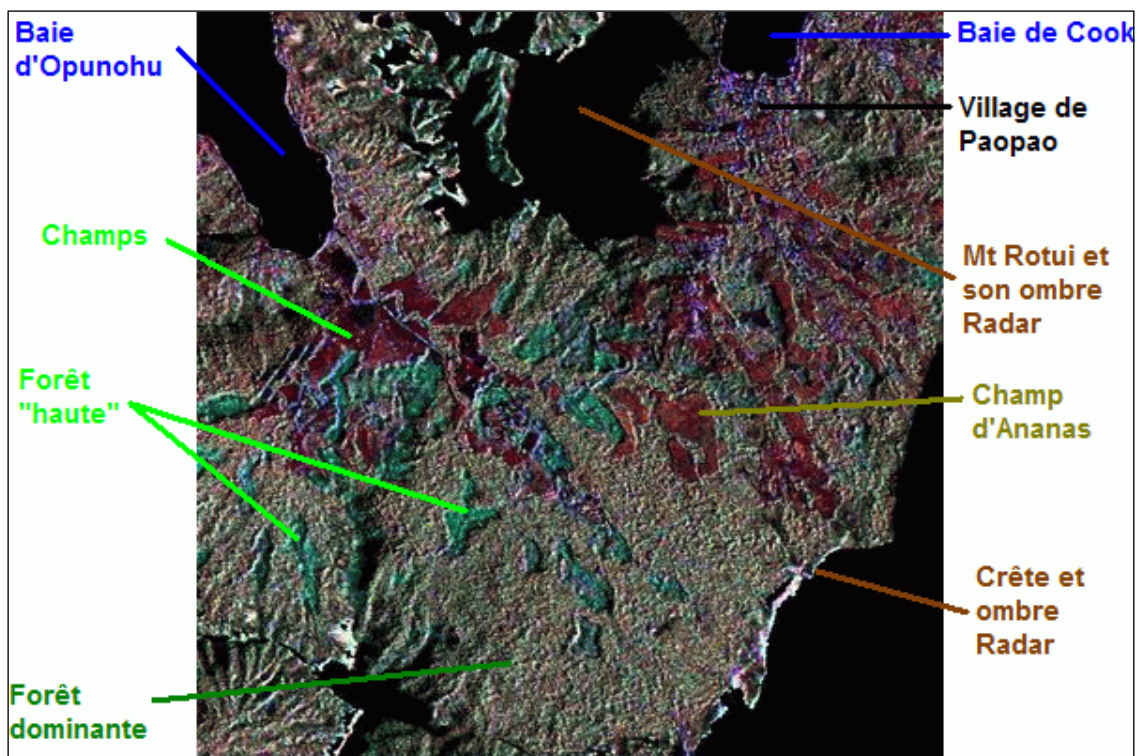


figure 22 Composite AirSAR (C_{VV} , L_{HV} , L_{HH}) - domaine d'Opunohu

Il faut rappeler ici le principe de l'imagerie radar : un signal actif est émis sur une porteuse de fréquences radar (C , L , ou P dans ce cas). Cette onde radar en heurtant la cible subit un certain nombre de changements de ses propriétés intrinsèques. Elle peut être absorbée, renvoyée de façon spéculaire (rebond), ou partiellement renvoyée avec dépolarisation éventuelle vers la source où elle est captée et enregistrée. Des informations sur la cible sont déduites de la mesure en amplitude et en phase du signal rétrodiffusé.

L'image radar est donc une image en niveau de gris, fortement bruitée, dont la valeur pour un pixel correspond à la quantité de signal rétrodiffusé pour une fréquence et une polarisation donnée. Pour bien comprendre la signification d'une image radar, il convient de choisir une représentation adaptée. En effet, les images radar, bien qu'intrinsèquement très riches en informations, présentent une apparence austère peu propice à l'interprétation néophyte.

¹² Modèle numérique d'élévation

Le composite RVB (C_{VV} , L_{HV} , L_{HH}) de la *figure 22* permet de faire ressortir les différents types d'occupation des sols.

- Les zones d'eau ou de sols nus ressortent en noir (réflexions spéculaires) : *i.e.* pas de signal radar rétrodiffusé vers le capteur.
- Les pâturages et prairies à végétation basse ressortent en rouge : prédominance de la bande C_{VV} due au signal rétrodiffusé par la rugosité de surface dans cette longueur d'onde.
- Les forêts hautes ressortent en vert : forte réponse de la bande L_{HV} , ce qui est pertinent avec une forte dépolarisation due aux branchages.
- La forêt dominante à *Hibiscus* (forêt moyenne) ressort en vert clair teinté de rouge, bleu, jaune : elle est caractérisée par une forte entropie.
- Les plantations d'ananas ont une couleur orangée caractéristique : forte réponse en C_{VV} et dans une moindre mesure en L_{HV} .
- Les zones anthropisées (zone urbaine) ressortent en violet : réponse dans les bandes C_{VV} et L_{HH} (double réflexions dues aux bâtiments)

Pour effectuer la classification supervisée, une campagne de vérité terrain a été conduite. Cette campagne a permis de proposer le jeu de classes suivant et son code couleur (*tableau 10*).

Forêt Haute	HF
Forêt Moyenne	MF
Végétation Basse	LV
Végétation Absente	NV
Cultures d'Ananas	PF
Zone Urbaine	UZ

tableau 10 Jeu de classes Opunohu - AirSAR

La *figure 23* illustre la signature normalisée des différentes classes sur les 4 bandes radar considérées, nous notons que chaque classe a une signature caractéristique.

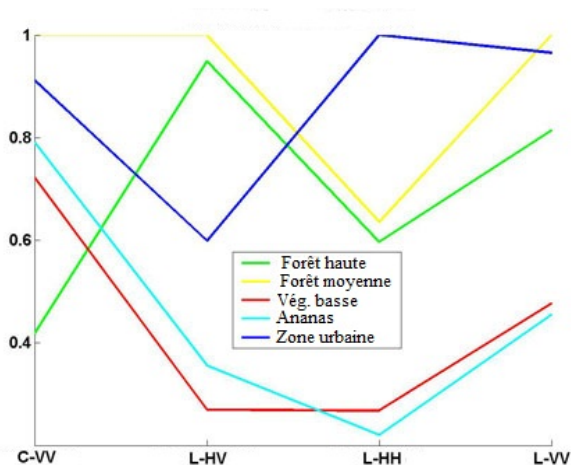


figure 23 Signature normalisée du jeu de classes

Cela est corroboré par un calcul de l'indice de séparabilité de Jeffries-Matusita représenté dans le *tableau 11* ci-après (plus cet indice est proche de 2, meilleure est la séparabilité statistique des paires de classes concernées).

Paire de classes		Indice de Jeffries-Matusita
Végétation basse	Champ d'ananas	1,26
Forêt haute	Forêt moyenne	1,75
Forêt moyenne	Zone urbaine	1,86
Forêt haute	Zone urbaine	1,90
Végétation basse	Zone urbaine	1,90
Champ d'ananas	Zone urbaine	1,91
Forêt moyenne	Champ d'ananas	1,99
Végétation basse	Forêt moyenne	1,99
Végétation basse	Forêt haute	2,00
Forêt haute	Champ d'ananas	2,00

tableau 11 Séparabilités de Jeffries-Matusita - Opunohu

Nous notons ici que les couples de classes les moins séparables sont les classes *végétation basse* et *champs d'ananas*, et les classes *forêt haute* et *forêt moyenne*. Ces deux cas s'expliquent par les similarités structurelles entre les classes concernées. Cependant, comme nous le constaterons sur les résultats de classification, ces couples ont une séparabilité suffisante pour qu'un algorithme de classification classique puisse les discriminer avec succès.

Un algorithme de classification supervisée par *maximum de vraisemblance* a été utilisé pour classifier l'image radar. La *figure 24* illustre la classification obtenue :

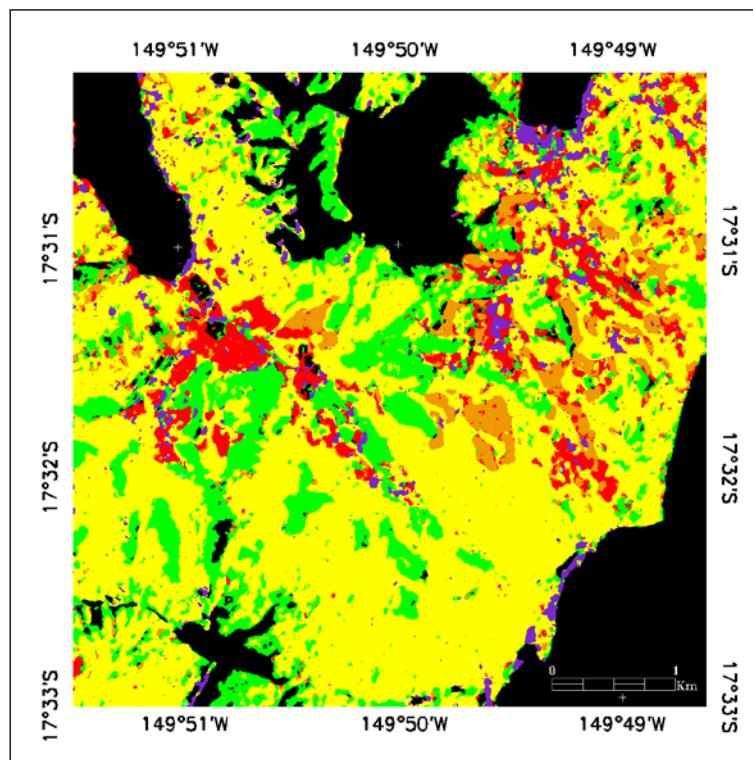


figure 24 Classification max. de vraisemblance AirSAR – Opunohu

La matrice de confusion du *tableau 12* nous aide à percevoir la qualité de la classification (*cf. chapitre III.A.2*) par comparaison entre les classes attribuées par l'algorithme de classification et la classe indiquée par la vérité terrain (pour chaque pixel de la base de vérification).

Classes	Vérité terrain				
	Zone urbaine	Forêt moyenne	Forêt haute	Champ d'ananas	végétation basse
Non-classifié	11,70%	0,00%	6,32%	0,00%	0,00%
Zone urbaine	51,33%	0,92%	17,12%	1,72%	15,52%
Forêt moyenne	23,87%	91,45%	3,50%	3,97%	3,41%
Forêt haute	9,83%	5,74%	72,29%	0,00%	0,27%
Champ d'ananas	0,47%	1,88%	0,33%	86,86%	9,55%
végétation basse	2,81%	0,00%	0,43%	7,45%	71,25%

Overall Accuracy 77,86% Kappa 0,7198

tableau 12 Matrice de confusion - Opunohu

La classe *zone urbaine* est très mal classifiée avec 51,33% de pixels bien classifiés. La classe la mieux classifiée est la classe *forêt moyenne* avec 91,45% de pixels bien classifiés. Les classes *végétation basse* et *forêt haute* obtiennent des performances honorables (essentiellement dégradées par une confusion avec la classe zone urbaine). La classe *champs d'ananas* obtient de bons résultats et se distingue en particulier de la classe *végétation basse*.

Malgré un pré-filtrage par un filtre de Lee, la classification est relativement bruitée, ce qui est dû au phénomène de chatoiement (*speckle*) de la donnée radar. Ce chatoiement peut être filtré grâce à des filtres adaptés ou intégrés dans les modèles de classification. Dans le cas d'un pré-filtrage comme c'est le cas ici, toute la difficulté résulte du fait que l'on doit trouver un compromis entre le filtrage et la dégradation des détails de l'image qu'il entraîne.

Cette étude préliminaire montre que la donnée radar en polarisations simples parallèles et orthogonales a des propriétés permettant de discriminer des classes de végétations simples. Bien sûr ces résultats demandent à être améliorés par l'utilisation d'algorithmes adaptés aux caractéristiques intrinsèques des données radar, par l'étude des propriétés polarimétriques de cette donnée et par une optimisation de la chaîne de traitement dans son ensemble.

Références : (Stoll et al. igarss2004; Stoll et al. arfp2004).

b. Un indice de végétation radar

Dans cette approche pour caractériser le couvert végétal polynésien, il est apparu intéressant d'étudier brièvement l'intérêt potentiel des indices de végétation. Il en existe beaucoup dans la littérature, l'un des plus classiques étant l'indice $NDVI^{13}$ (Eq.2) :

$$NDVI = 100 \times \left(\frac{IR - R}{IR + R} + 1 \right) \quad Eq.2$$

L'indice $NDVI$ a été calculé à partir des bandes rouge et infrarouge de la donnée MASTER (Modis-ASTER) sur le site d'étude d'Opunohu (*figure 25 - gauche*).

Les statistiques de l'indice $NDVI$ par classe (*tableau 13*) montrent que cet indice ne permet pas de discriminer les classes de végétation denses des classes de végétation peu denses. En effet, l'indice $NDVI$ donne une indication sur la présence ou non de végétation et se révèle donc peu pertinent pour étudier la végétation de nos îles où la végétation est omniprésente.

Or, certaines bandes radar sont intéressantes pour distinguer les zones à fort couvert végétal de celles dépourvues de végétation. J'ai donc proposé l'indice de végétation radar $SARvi$ suivant (Eq. 3.) :

¹³ Tucker, C. J. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of the Environment, 1979, 8, 127-150.

$$SARvi = 100 \times \left(\frac{L_{HV} - C_{VV}}{L_{HV} + C_{VV}} + 1 \right) \quad Eq.3$$

Cet indice a pour avantage de n'être basé que sur les deux seules bandes radar C_{VV} et L_{HV} , la figure 25 - droite illustre celui-ci sur la zone d'Opunohu.

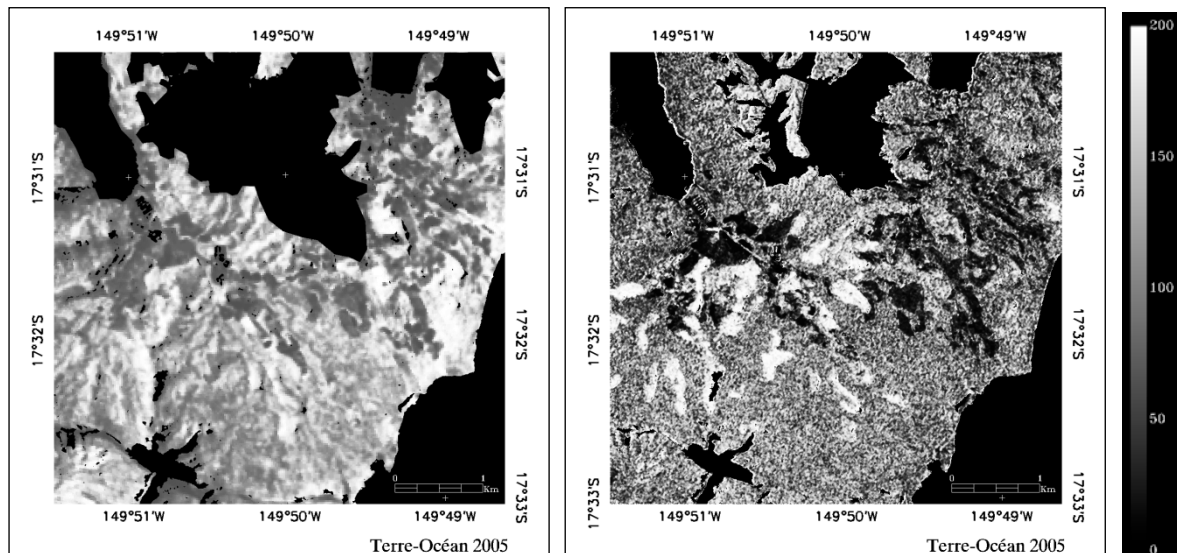


figure 25 Indices NDVI (gauche) et SARvi (droite) - Opunohu

La comparaison visuelle des deux images (figure 25) ne permet pas de conclure, ne serait-ce qu'en raison de la différence de résolution et de dynamique des images. En revanche, l'analyse des statistiques de chaque classe (tableau 13) montre que contrairement à l'indice NDVI, l'indice SARvi permet une discrimination des classes de végétation de notre site d'étude.

	NDVI		SARvi	
	Moyenne	Ecart type	Moyenne	Ecart type
Pas de végétation	132,7	33,0	40,9	18,0
Végétation basse	143,8	7,3	58,8	12,8
Forêt moyenne	154,9	23,5	100,2	15,0
Forêt haute	150,2	25,3	139,8	24,3

tableau 13 Valeur des indices NDVI et SARvi par classe - Opunohu

Quand l'indice NDVI varie peu en fonction de la densité de végétation concernée, l'indice SARvi croît de façon continue : de 40 pour les zones dépourvues de végétation à 140 pour les zones à forte densité de végétation. Cette propriété en fait un indice très utile pour détecter les différentes densités de végétation.

Pour appliquer le protocole de Kyoto, les gouvernements sont demandeurs de données chiffrées de biomasse, cet indice SARvi pourrait permettre de les fournir sur des zones tropicales difficiles d'accès. Pour ce faire, il faudrait établir une relation entre l'indice SARvi et la biomasse, et par conséquent effectuer une mission terrain pour relever la biomasse *in situ*. Enfin, pour que cette relation soit la plus précise possible, il faudra travailler sur les aspects de la calibration de la donnée radar, de la réduction des effets de pentes etc.

Références : (Stoll igarss2005; Stoll et al. perf2006).

c. Étude comparative des bandes radar simples

La donnée radar polarimétrique JPL-AirSAR est une donnée très riche comportant trois fréquences radar (C , L , P) dont les bandes L et P sont en mode polarimétrique. Cependant, son coût est prohibitif car elle nécessite une campagne de prise de vue aéroportée avec la prise en charge des frais inhérents à la mobilisation d'un avion et de son équipage.

De ce fait, le choix se porte plus souvent sur une acquisition de données satellite radar, plus rapides à obtenir et moins onéreuses. Toutefois, il n'existe pas de satellite radar ayant les mêmes capacités en termes de mode de prises de vue que la donnée AirSAR, ceci pour des raisons liées à la technologie satellitaire.

Le chercheur ou l'utilisateur de données satellitaires devra donc souvent se contenter d'un nombre limité de fréquences et de polarisations (le cas le plus fréquent est une seule fréquence radar avec une double polarisation). Devant la diversité de l'offre des différents satellites radar, il est souvent difficile de faire un choix. Quelles fréquences choisir ? Quelles polarisations choisir ?

Puisque nous disposons de données AirSAR polarimétriques avec 3 fréquences radar, une comparaison des différentes combinaisons de 3 bandes en polarisations linéaires a été conduite parmi les bandes C , L et P , soient : C_{VV} , L_{HH} , L_{HV} , L_{VV} , P_{HH} , P_{HV} , P_{VV} .

La figure 26 illustre la précision moyenne (MPA^{14}) des classifications SVM de chaque triplet radar.

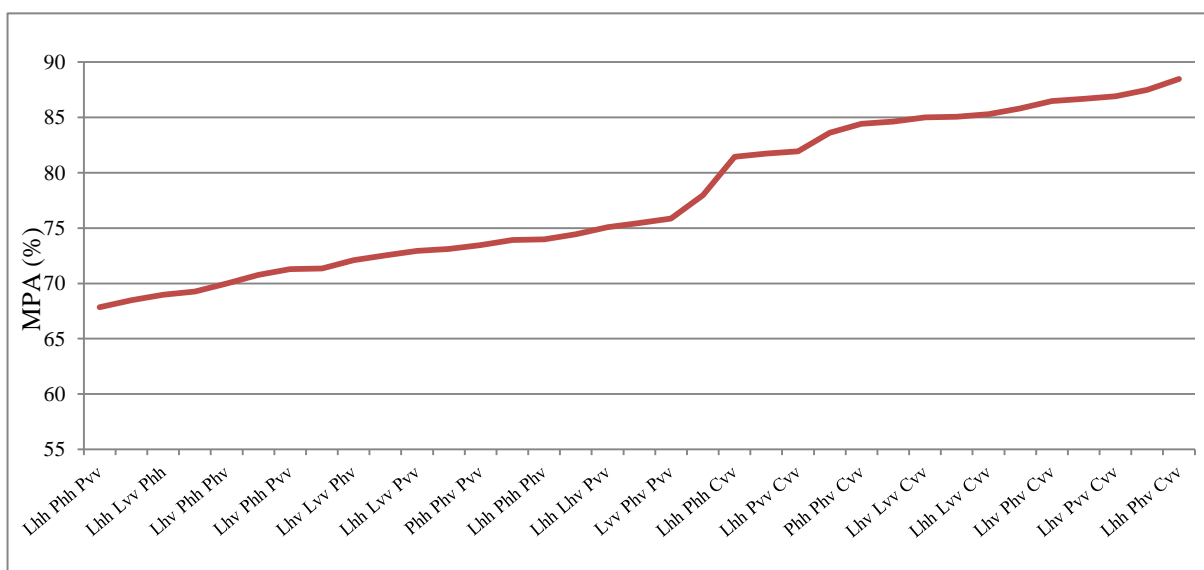


figure 26 Précision des classifications SVM de triplet de bandes radar

La meilleure performance est obtenue pour la combinaison L_{HH} , P_{HV} , C_{VV} avec seulement 2% de moins que l'ensemble des polarisations simples des bandes C , L et P , cela montre qu'un choix judicieux permet une économie substantielle de données et de temps de traitement.

Notons que la bande C_{VV} apporte un gain de 4 à 8% de la précision globale. Cela est lié à sa faible longueur d'onde qui la rend plus sensible aux textures plus fines de la canopée, démontrant ainsi son intérêt pour discriminer les forêts denses (plus spécifiquement le *Pinus* et le *Falcata*).

Références : (Lardeux et al. grsl2010).

¹⁴ Mean Producer Accuracy

2. Donnée radar polarimétrique

Ces premières études ont mis en évidence l'intérêt de l'utilisation de la donnée JPL-AirSAR (par classification des bandes radar en polarisations simples C_{VV} , L_{HV} , L_{HH} , L_{VV}) dans le contexte polynésien.

Une étude a été initiée afin d'explorer et valider la potentialité de la donnée radar polarimétrique en exploitant la polarisation complète des bandes L et P . Le site d'étude choisi est l'île de Tubuai dans l'archipel des Australes.

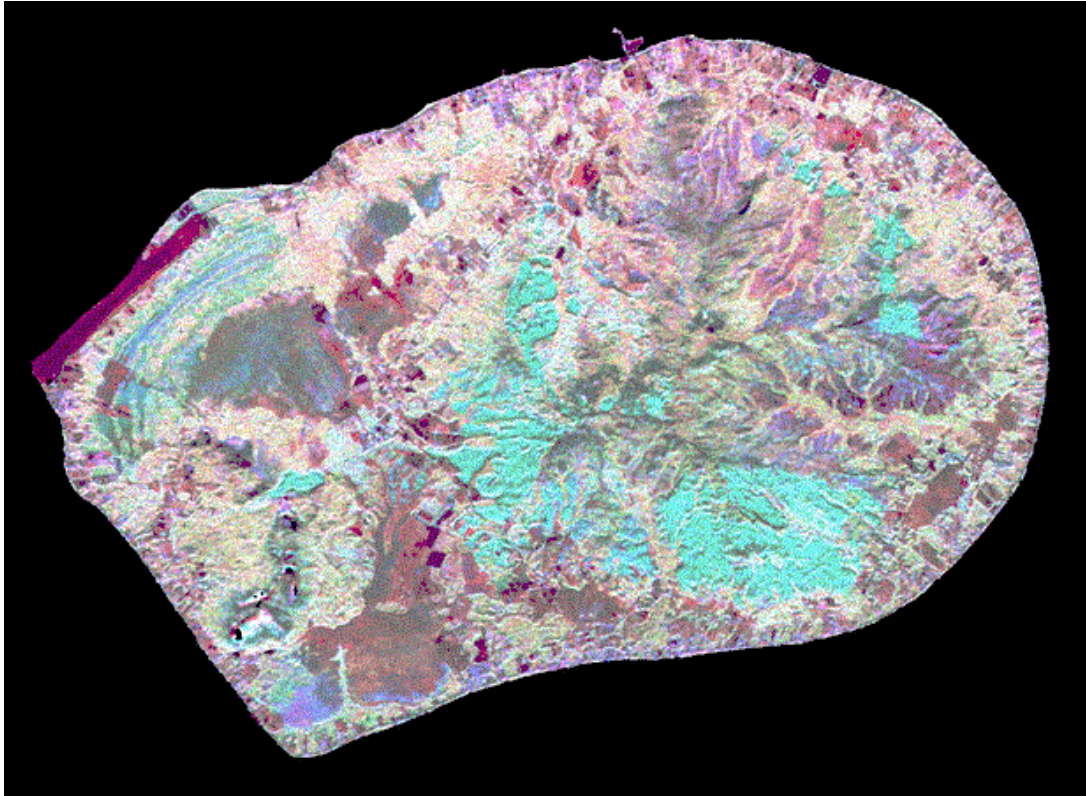


figure 27 Composite JPL-AirSAR (C_{VV} , L_{HV} , L_{HH}) Tubuai

L'île de Tubuai a une configuration relativement simple : deux petits massifs montagneux, deux grands marais à l'ouest, des cultures variées, des forêts de *Pinus*, de *Falcata*, des landes à fougères *Anuhe*, ...

Sur l'image composite de la *figure 27*, nous remarquons les mêmes types de signatures que dans le cas de Moorea, avec une particularité cependant : la présence de deux marais. Ceux-ci ont une signature très hétérogène due aux différences de taux d'humidité ainsi qu'aux différents types et concentrations de roseaux que l'on y trouve.

a. Apport des indices polarimétriques

La matrice de cohérence T est dérivée de la donnée *Multi-Look Complexe* JPL-AirSAR, elle est définie à partir du vecteur de rétrodiffusion de la cible exprimé dans la base de Pauli. Il sera établi au *chapitre III.B.3* qu'il est possible de classifier la donnée radar polarimétrique à partir des 9 éléments de la matrice de cohérence, laquelle contient toute l'information polarimétrique.

De cette matrice de cohérence, il est possible d'extraire un certain nombre d'indices polarimétriques permettant de mettre en avant des propriétés particulières de la cible. Devant le nombre conséquent d'indices disponibles dans la littérature se pose la question de savoir

dans quelle mesure l'ajout de tout ou partie de ces indices en plus des éléments de la matrice de cohérence permettrait d'améliorer la précision des classifications.

Avant de procéder à une analyse exhaustive de tous les indices polarimétriques, une première étude a été menée pour vérifier si l'ajout d'une partie d'entre eux permettait d'améliorer les précisions des classifications.

Pour cela, les classifications effectuées sur les 9 éléments de la matrice de cohérence ont été comparées avec celles effectuées avec l'ajout d'une partie des indices polarimétriques, soit 17 bandes au total. Se référer à (Lardeux *et al.* 2006) pour la description des indices concernés.

MPA	SVM mat. cohérence	SVM mat. cohérence + 17 IP
Bande L	62,60%	69,50%
Bande P	61,64%	72,69%

tableau 14 Pertinence de l'ajout d'indices polarimétriques

Le tableau 14 reporte les précisions MPA obtenues pour les bandes L et P. Il met en évidence que pour la bande L comme pour la bande P, les résultats sont améliorés sensiblement lorsque l'on ajoute les 17 indices polarimétriques aux 9 éléments de la matrice de cohérence.

Quelle influence peut avoir l'utilisation d'une ou plusieurs bandes (fréquences de la porteuse radar) sur le résultat des classifications ? Pour répondre à cette question, l'algorithme de classification SVM a été appliqué sur les éléments de la matrice de cohérence additionnés des 17 indices polarimétriques dans les cas suivants :

- La bande L seule (9+17 primitives)
- La bande P seule (9+17 primitives)
- Les bandes L et P conjointement (9+17+9+17 primitives)
- Les bandes L et P additionnées de la bande C_{VV} (9+17+9+17+1 primitives)

SVM mat. cohérence +			
IP (bande L)	IP (Bande P)	IP (bandes L+P)	IP (bandes L+P) + C_{VV}
69,50%	72,69%	84,31%	93%

tableau 15 Pertinence de l'ajout de fréquences radar

Les résultats du tableau 15 sont éloquent, l'utilisation de toutes les bandes de fréquences radar ainsi que de leurs indices polarimétriques correspondants permet d'améliorer les résultats de classification (MPA) de manière significative. Il semble *a priori* que plus on injecte d'information, plus les résultats de classification s'améliorent.

Cependant, le nombre de bandes à classifier devient important et cela augmente considérablement les temps de calcul. A la lumière de cet élément, la question se pose de savoir si l'ajout de la totalité des indices polarimétriques est nécessaire. Pour y répondre, le tableau 16 présente une comparaison des résultats précédents avec ceux de la classification des trois seules bandes C_{VV} , L_{VV} , P_{HV} .

Bandes	Mat. cohérence(L+P) + IP(L+P) + C_{VV}	C_{VV} , L_{VV} , P_{HV}
MPA	93,00%	81,69%

tableau 16 Comparaison avec le cas C_{VV} , L_{VV} , P_{HV} .

Même si les résultats de classification sont moins bons que dans le cas où l'on utilise toutes les bandes ainsi que leurs indices polarimétriques, l'utilisation des trois bandes C_{VV} , L_{VV} , P_{HV} seules permettent d'obtenir de très bons résultats avec un temps de calcul considérablement réduit puisque l'on passe de 35 bandes à 3.

Le *tableau 17* synthétise les résultats précédents en spécifiant le nombre de bandes correspondant.

	Mat. Coh. + IP (L)	Mat. Coh. + IP (P)	Mat. Coh. + IP (L+P)	Mat. Coh. + IP (L+P) + Cvv	C_{VV} , L_{VV} , P_{HV}
MPA	69,50%	72,69%	84,31%	93%	81,69%
nbre	26	26	52	53	3

tableau 17 Comparaison globale

En conclusion, l'utilisation de la polarimétrie peut parfois être supplantée par un choix judicieux de bandes et polarisations aux caractéristiques complémentaires, et ceci avec une complexité algorithmique moindre.

Voici la classification SVM de la donnée AirSAR de Tubuai utilisant les bandes C_{VV} , L et P avec tous les indices polarimétriques correspondants (*figure 28*).

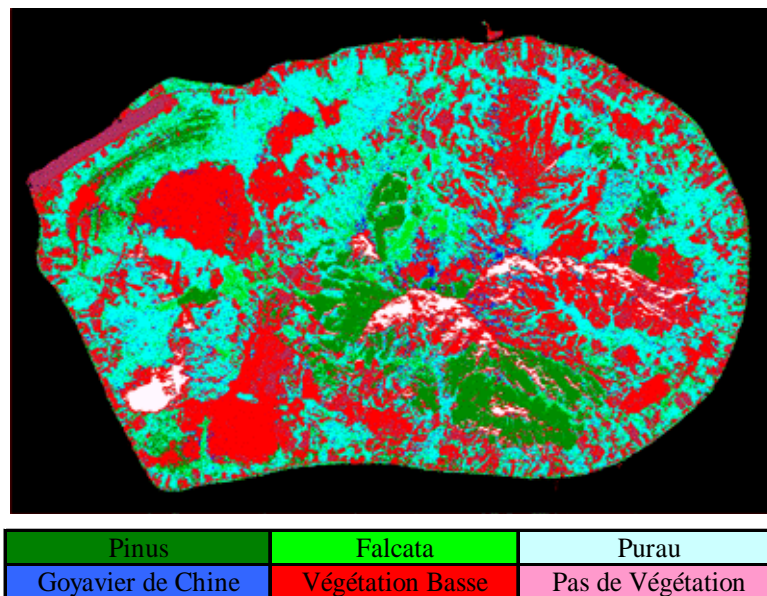


figure 28 Classification de Tubuai

De façon générale, nous verrons dans ce document que même si l'idée d'ajouter des couches d'informations est souvent tentante, cela augmente le temps de calcul ainsi que la complexité du problème. Ainsi, il faut toujours s'interroger sur la pertinence de la donnée que l'on veut ajouter, et vérifier que son utilisation n'entraîne pas une dégradation des performances des algorithmes.

Références : (Lardeux et al. igarss2006).

b. Étude comparative des modes polarimétriques

L'intérêt de certains de ces indices polarimétriques pour la classification du couvert végétal polynésien ayant été validé au chapitre précédent, il est intéressant d'étudier l'utilisation des modes polarimétriques partiels disponibles sur certains satellites.

En effet, avec de la donnée polarimétrique JPL-AirSAR à notre disposition, il était possible de simuler ces modes polarimétriques partiels.

Nous avons donc comparé la polarimétrie complète avec la polarimétrie partielle disponible sur les deux capteurs radar ENVISAT/ASAR et ALOS/PALSAR ainsi que le mode expérimental $\pi/4$. Pour chaque capteur, nous avons classifié l'intégralité des indices polarimétriques qu'il était possible d'extraire.

Le *tableau 18* présente la liste des 54 indices polarimétriques extraits de la matrice de cohérence en mode polarimétrie complète pour chacune des bandes L et P .

Description	Designation	Nombre d'éléments
Éléments de la matrice de cohérence (Re et Im)	T_{ij} ($i, j = 1, 2, 3$)	9
Intensités linéaires et circulaires	$I_{hh}, I_{vv}, I_{gg}, I_{dd}$	4
Span	$S = I_{hh}^2 + 2 I_{hv}^2 + I_{vv}^2$	1
Rapports d'intensités linéaires et circulaires	$I_{hv}/I_{vv}, I_{hv}/I_{hh}, I_{hh}/I_{vv}$ $I_{dd}/I_{gd}, I_{gg}/I_{gd}, I_{gg}/I_{dd}$	6
Degrés de cohérence entre les polarisations linéaires (Module et Phase)	$\rho_{hh-vv}, \rho_{hh-hv}, \rho_{vv-hv}$	6
Degrés de cohérence entre les polarisations circulaires (Module et Phase)	$\rho_{gg-dd}, \rho_{gg-gd}, \rho_{dd-gd}$	6
Extrema d'intensité de polarisation	$I_{min}, I_{max} / I_{min}$	2
Extrema du degré de polarisation	$dp_{min}, dp_{max}, \Delta(dp)$	3
Coefficients de texture d'intensités linéaires et circulaires	$\Gamma x_{hh}, \Gamma x_{hv}, \Gamma x_{vv}$ $\Gamma x_{gg}, \Gamma x_{gd}, \Gamma x_{dg}$	6
Paramètres d'Euler	m, γ, ψ, v, τ	5
Entropie, anisotropie, paramètre α	$H/a/\alpha$	3
Paramètres de la décomposition de Freeman	P_s, P_d, P_v	3

tableau 18 Les 54 indices polarimétriques

Le mode de polarisation alternée (AP) du capteur ASAR présent sur le satellite ENVISAT consiste en l'émission d'une polarisation linéaire (H) et la réception dans deux polarisations linéaires (HH et HV , VV et HV , ou HH et VV). Le *tableau 19* liste les 6 indices polarimétriques de ce mode. Les indices w et x font référence aux polarisations linéaires horizontales (H), ou verticales (V).

$SV_{ASAR DP}$	Nombre éléments
I_{wx}, I_{wx}	2
$SPAN = S_{wx} ^2 + S_{wx} ^2$	1
C_{V-wx}, C_{V-wx}	2
$\frac{I_{wx}}{I_{wx}}$	1

tableau 19 Indices polarimétriques du mode polarimétrie alternée (ASAR).

Le mode polarimétrie duale (DP) du capteur PALSAR du satellite ALOS consiste en l'émission d'une onde polarisée linéairement et en la réception de l'onde rétrodiffusée dans deux polarisations linéaires. Le résultat est un couple d'intensité de rétrodiffusion de deux

polarisations (HH/HV , VV/HV ou HH/VV) additionné de la différence de phase entre les deux polarisations. Les 13 indices polarimétriques associés sont décrits dans le *tableau 20* suivant.

$SV_{\text{PALSAR-DP}}$	Nombre éléments
C2	4
SPAN = $ S_{sp} ^2 + S_{sp} ^2$	1
C_{v-pq} , C_{v-pq}	2
$\frac{I_{pq}}{I_{pq}}$	1
$ \rho_{pq-pqv} $, φ_{pq-pq}	2
H/a/ α (mode DP)	3

tableau 20 Indices polarimétriques du mode polarimétrie duale (PALSAR)

Le mode expérimental $\pi/4$ consiste en une polarisation d'envoi du signal circulaire ou orientée à 45° , et en une polarisation horizontale ou verticale à la réception. Grâce à des propriétés de symétrie sur les cibles naturelles, il est possible selon certaines hypothèses de reconstruire l'information de polarisation complète. Ce type de données a montré son efficacité sur des cibles distribuées comme les données d'occupation des sols.

Le *tableau 21* suivant liste les 31 indices polarimétriques associés au mode $\pi/4$.

SV_{ENVISAT}	Nombre éléments
T elements	7
I_{HH} , I_{HV} , I_{VV} , I_{LL} , I_{LR}	5
$SPAN = S_{HH} ^2 + 2 S_{HV} ^2 + S_{VV} ^2$	1
C_{v-HH} , C_{v-HV} , C_{v-VV} , C_{v-LL} , C_{v-LR} ,	5
$\frac{I_{H_v}, I_{H_v}, I_{H_H}, I_{L_L}}{I_{H_H}, I_{V_v}, I_{V_v}, I_{L_R}}$	4
$ \rho_{H_{HV}} $, $ \rho_{L_{LL}} $	2
P_{\min}	1
Deg_{\min}	1
m and γ Euler parameters :	2
H/A/ α	3

tableau 21 Indices polarimétriques du mode " $\pi/4$ "

Le *tableau 22* résume la comparaison des résultats de classifications *SVM* (*MPA*) du mode polarimétrique complet (*FP*) et des modes polarimétriques partiels $\pi/4$, polarimétrie duale et alternée.

Le mode polarimétrique complet obtient les meilleures performances, mais nous remarquons que certaines polarisations partielles permettent d'obtenir des précisions quasiment équivalentes, surtout lorsque l'on utilise plusieurs fréquences.

Notons aussi que le mode dual polarisation (*DP*) permet généralement d'atteindre de meilleures précisions que le mode de polarisation alternée (*AP*).

	FP	Polarimétrie duale			$\pi/4$	Polarimétrie duale			Polarimétrie alternée			Polarimétrie alternée		
		HH/HV HH/NV	VV/HV HH/NV	HH/HV VV/HV		HH/NV	HH/HV	VV/HV	HH/HV HH/NV	VV/HV HH/NV	HH/HV VV/HV	HH/NV	HH/HV	VV/HV
L	90	83	84	80	78	74	72	69	72	72	72	62	64	62
P	90	84	84	83	78	72	71	73	74	75	74	58	64	68
L+C	98	96	96	95	94	92	92	89	92	92	92	86	88	86
P+C	97	95	96	94	93	91	89	90	91	91	91	83	85	87
L+P	98	97	97	97	93	90	87	88	90	91	90	80	80	82
L+P+C	99	99	99	99	98	98	96	96	97	97	97	94	94	95

tableau 22 Comparaison des modes polarimétriques complets et partiels (%)

Références : (Lardeux et al. polinsar2007; Lardeux et al. envisat2007; Lardeux et al. alos2007; Lardeux et al. grsl2010)

c. Contribution relative des indices polarimétriques

Les résultats précédents montrent que même si les modes de polarisation partielle sont une solution intéressante pour obtenir de bonnes performances avec un nombre réduits de bandes, l'utilisation de la polarimétrie complète et des indices polarimétriques associés permettent d'obtenir les meilleures précisions.

Toutefois, augmenter le nombre d'indices à classifier entraîne un accroissement du temps de calcul et des problèmes se posent pour le traitement d'un grand nombre de bandes. De plus, nous avons vu que l'ajout de couches d'informations dans un algorithme de classification n'améliore pas systématiquement les précisions, celle-ci peut décroître si l'information n'est pas pertinente, trop corrélée, ou bruitée.

Nous devons donc nous intéresser à la pertinence individuelle des multiples indices du *tableau 18*. En effet, il semble logique de penser que les 54 indices ne contribuent pas tous de la même manière à améliorer la qualité d'une classification. Nous allons essayer de déterminer quels indices apportent le gain le plus important en termes de discrimination des classes de végétation présentes sur notre site d'étude.

Cependant l'analyse exhaustive de tous les cas possibles n'est pas envisageable, la méthode d'analyse de Greedy est donc appliquée afin de réduire le temps de traitement et déterminer ainsi un classement des indices les plus discriminants.

La méthode itérative de Greedy-backward permet de déterminer les indices les moins significatifs :

- Cette procédure commence par la classification *SVM* obtenue avec les 54 vecteurs initiaux.
- Ensuite, nous effectuons une classification *SVM* en calculant toutes les classifications correspondantes à toutes les combinaisons de 53 vecteurs parmi 54, la combinaison gardée est celle qui obtient la meilleure précision.

- Puis parmi les 53 vecteurs de cette combinaison gagnante, nous cherchons les 52 vecteurs qui obtiennent la meilleure précision.
- Cette procédure est répétée de proche en proche jusqu'à ce qu'il ne reste que 3 vecteurs. Nous notons à chaque étape la précision atteinte par la meilleure combinaison.

Et la méthode de Greedy-forward est utilisée pour déterminer les indices les plus significatifs :

- Nous commençons par effectuer toutes les classifications de 3 indices parmi les 54.
- Ensuite, nous gardons celle qui obtient la meilleure précision.
- Puis, nous ajoutons un indice parmi les 51 restants, nous choisissons celui qui obtient la meilleure précision.
- De proche en proche, nous ajoutons à chaque fois le vecteur le plus profitable à la classification.

La *figure 29* montre la précision des classifications obtenues par rapport au nombre de primitives du vecteur, ceci pour les bandes *L*, *P* et les trois bandes *L*, *P* et *C* conjointement.

Les deux bandes *L* et *P* ont un comportement similaire : un large plateau est atteint lorsque le nombre de primitives est compris entre 20 et 45. Il est à noter aussi une chute des performances lorsque le nombre de primitives dépasse 48 pour les deux bandes, ce qui tend à montrer que certains indices polarimétriques introduisent de la confusion dans l'algorithme de classification.

Une analyse des indices polarimétriques concernés montre que 7 indices en particulier font partie des indices les moins discriminants pour les deux bandes *L* et *P*. Il s'agit de 4 paramètres d'Euler ν , ψ , γ , τ et les trois phases de degré de cohérence φ_{ll-rr} , φ_{rr-lr} , et φ_{hh-hv} .

Ces résultats ne sont pas surprenants, puisque les paramètres d'Euler ont été définis pour la caractérisation de cibles déterministes en polarisation complète (or ceci n'est pas le cas sur le milieu naturel), les trois phases de degré de cohérence sont d'ailleurs généralement utilisées en milieu urbain.

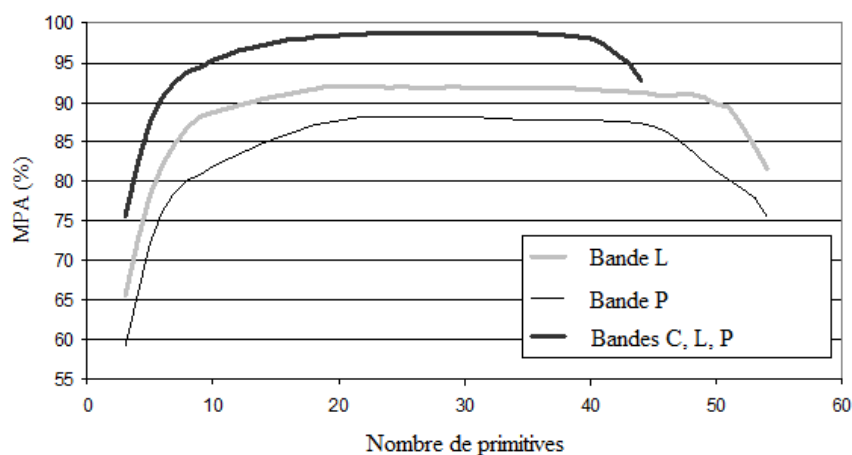


figure 29 Contribution des indices polarimétriques par bandes.

(Lardeux thesis2008) propose un classement des 54 indices polarimétriques par rapport à leur contribution pour la classification du couvert végétal de Tubuai.

Références : (Frison et al. igarss2008; Lardeux et al. spie2008; Lardeux et al. tgrs2009)

C. DONNEES ANCILLAIRES

Pour améliorer les précisions des classifications, l'utilisation de certaines données ancillaires, conjointement aux données optiques et radar, est envisagée. Ces données apportent une information complémentaire permettant d'expliquer la spatialisation des espèces. Leur utilisation va permettre de lever des ambiguïtés sur certaines classes.

1. Descripteurs biophysiques

De façon générale et comme présenté dans l'introduction, la répartition spatiale des différents couverts végétaux est naturellement contrainte par un certain nombre de facteurs comme l'altitude, la ressource en eau, l'insolation, etc.

La maîtrise de ces facteurs est d'une importance fondamentale pour appréhender la dynamique de la végétation et donc sa répartition actuelle.

a. Facteurs abiotiques

Voici la liste des facteurs abiotiques les plus directement accessibles qui ont été utilisés dans nos études :

- L'élévation : elle influe directement sur la température de l'air et donc l'évapotranspiration, la fixation et la décomposition du carbone, la productivité des plantes et leur mortalité.
- La pente : elle a une influence sur la vitesse d'écoulement de l'eau, l'érosion du sol, les effets mécaniques sur l'implantation des racines des plantes, etc.
- L'exposition : aussi appelé « aspect », elle a une influence sur l'évapotranspiration et l'insolation.
- L'indice topographique composé (*CTI*) : il quantifie le drainage de l'eau par la microtopographie et explique la géomorphologie.
- L'exposition aux vents dominants : il permet d'intégrer le phénomène de la côte sud-est plus humide et la côte nord-ouest plus sèche sur les îles hautes polynésiennes.
- Les précipitations : les cartes annuelles de précipitations fournies par les services de Météo-France en Polynésie française donnent une information sur la répartition spatiale de l'apport direct en eau.
- L'insolation annuelle potentielle (*API*) : elle est proportionnelle à l'énergie radiative reçue par le sol.

La plupart des facteurs abiotiques précédents (à l'exception des précipitations) sont calculés à partir du modèle numérique d'élévation (*MNE*).

b. Facteurs biotiques

Le principal facteur biotique utilisé est une carte du couvert végétal dominant pour des études sur la végétation de sous-bois, les espèces envahissantes ou rares. En effet, la couche supérieure de la végétation peut affecter la température, la lumière, l'apport en eau et la quantité de nutriments disponible en sous-bois.

En conclusion, tous ces descripteurs biophysiques ne sont pas à proprement parler des données de télédétection satellitaire. Cependant, en apportant une information sur les conditions de vie des plantes, ils vont contribuer à l'amélioration des précisions de nos classifications.

Parmi ces descripteurs biophysiques, certains sont plus ou moins pertinents, d'autres peuvent être redondants. Une mesure de corrélation et une étude *a posteriori* de l'influence de chaque descripteur permet de ne conserver que les plus utiles.

2. Application à la modélisation de la niche écologique de *Miconia*

On l'a vu dans le cas des cocotiers, le facteur le plus contraignant de la télédétection satellitaire est que seule la couche supérieure de la végétation (la canopée) est détectable puisque celle-ci masque les couches inférieures.

De la même manière, le problème se pose dans le cas du *Miconia calvescens*, espèce fortement envahissante en Polynésie française (principalement sur Tahiti, Moorea, et Raiatea). En effet, le *Miconia* pousse principalement en sous-bois, sur les fortes pentes et sort peu en canopée à l'exception des zones où une dégradation de la couche supérieure intervient (en cas de cyclone par exemple), auquel cas cette espèce prend une forme arbustive. Cette caractéristique le rend par conséquent indétectable dans la majorité des cas.

Dans les rares zones mono-spécifiques de *Miconia*, s'agissant de la donnée optique, des essais ont montré que sa signature n'est pas discriminante par rapport à celle des forêts dans lequel il pousse. De plus, le *Miconia* pousse préférentiellement dans les zones à l'ombre le matin (heure de passage des satellites imageurs en Polynésie).

L'utilisation de donnée radar n'a pas non plus été concluante, que ce soit sur de la donnée polarimétrique JPL-AirSAR ou de la donnée TerraSAR-X à cause de ses caractéristiques structurales trop proches de la végétation environnante et des effets de pentes.

Les techniques de télédétection mono-sources directes sont donc peu adaptées pour détecter le *Miconia*.

La première solution envisagée est inspirée des techniques de modélisation de niches écologiques. Ces techniques sont couramment employées par les biologistes pour cartographier l'habitat potentiel de certaines plantes. Cette modélisation part du postulat que les plantes poussent uniquement sur certains milieux décrits par des caractéristiques topographiques précises. Par exemple, en Polynésie française, les fougères *Anuhe* sont principalement présentes sur les crêtes sommitales, les *Mape* (*Inocarpus fagifer*) poussent dans les fonds de vallées humides, ...

Nous avons comparé les résultats du modèle *GARP*¹⁵, avec une classification *SVM* binaire (2 classes : présence ou absence du *Miconia*). Ces deux techniques sont basées sur les mêmes variables géomorphologiques que sont : l'élévation, la pente, l'aspect, l'exposition aux vents, et un indice d'humidité.

Une mission terrain a permis de récolter des données de vérité terrain notant la présence ou l'absence du *Miconia* et ainsi permettre l'apprentissage et la validation des deux algorithmes.

Les résultats montrent que l'algorithme *SVM* apporte la meilleure précision avec 92.5% de précision globale (*Overall Accuracy*) contre 72.5% pour le modèle *GARP*. Les cartographies obtenues avec l'algorithme *SVM* sont moins homogènes et semblent plus sensibles à la microtopographie, ce qui pourrait expliquer ses meilleurs résultats.

La première cartographie précise de l'habitat potentiel du *Miconia Calvescens* sur l'île de Moorea a été produite (*figure 30*). La surface de l'invasion du *Miconia* selon les résultats de l'algorithme *SVM* est de 25% (35 km²).

¹⁵ Genetic Algorithm for Rule-Set Production

Une classification *SVM* de descripteurs biophysiques simples permet donc de cartographier l'habitat potentiel d'une espèce envahissante de sous-bois comme le *Miconia Calvescens*. Cela atteste de l'intérêt des méthodes de détection indirecte.

Références : (Pouteau, Meyer et al. isrse2011).

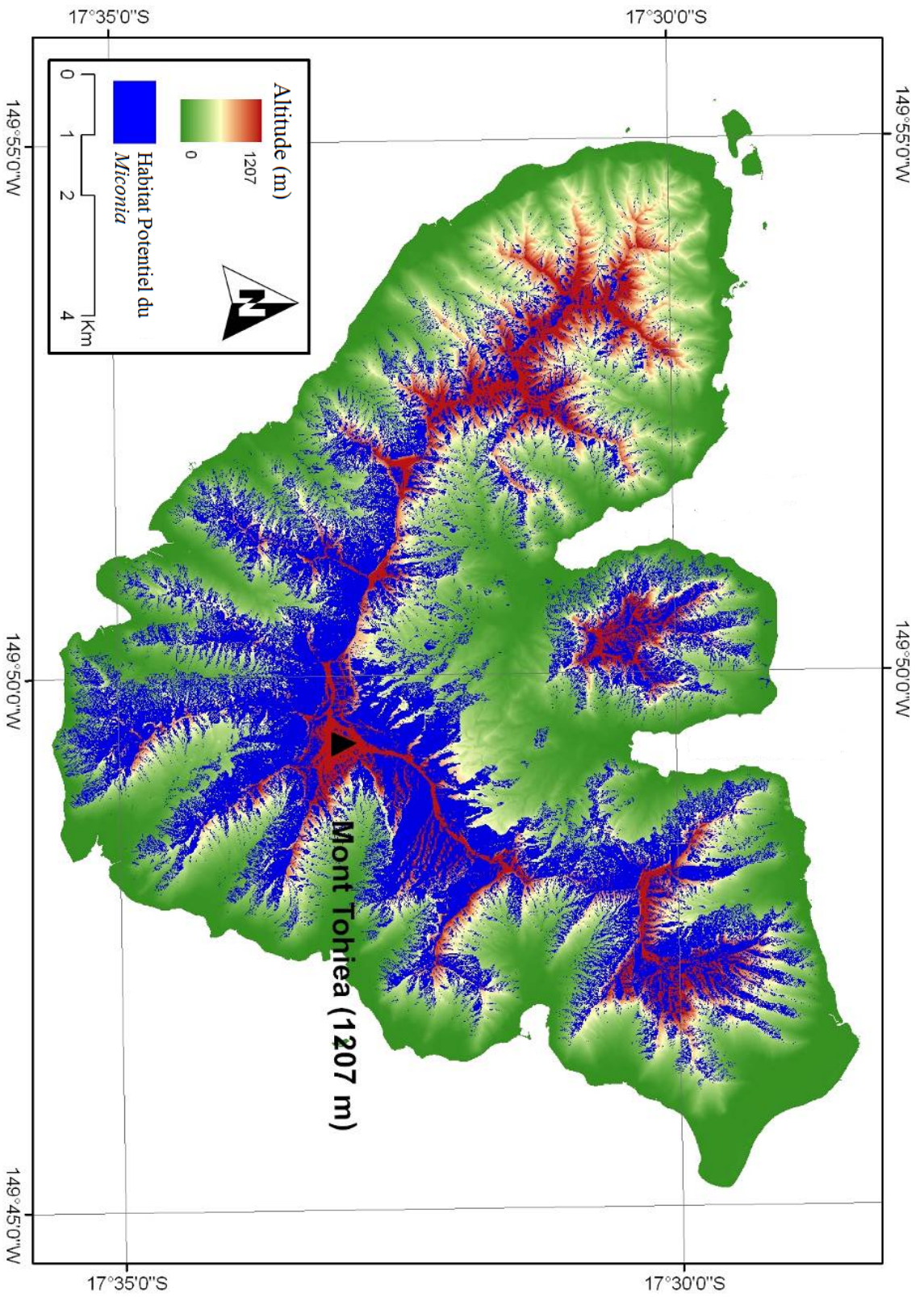


figure 30

Cartographie SVM de l'habitat potentiel du *Miconia* – Moorea

III. LES ALGORITHMES DE CLASSIFICATION

A. PREAMBULE

Intéressons-nous maintenant aux outils, c'est-à-dire aux pinceaux, couteaux, fusains du peintre. Pour le télédéacteur il s'agit des algorithmes.

Ceux-ci sont des programmes informatiques permettant de résoudre un certain nombre d'opérations liées à l'imagerie satellitaire. Ces opérations peuvent être : la correction radiométrique, la correction géométrique, le géo-référencement des images, le changement de projection, le pan-affinage (pan-sharpening), le mosaïquage, le filtrage, la segmentation, la classification, etc.

L'aspect algorithmique des études en télédétection est d'une importance majeure, il existe, pour chaque type d'opération, de nombreux algorithmes basés sur des théories mathématiques diverses.

Ma démarche consiste à acquérir un niveau d'expertise me permettant de sélectionner et d'utiliser un ou plusieurs algorithmes adaptés à une problématique donnée plutôt que d'en développer de nouveaux.

De plus, les images satellites sont généralement livrées avec les prétraitements de base tels que la correction radiométrique, la correction géométrique, le géo-référencement et le mosaïquage. Nous ne nous attarderons pas sur ceux-ci, ils sont en général de qualité satisfaisante (à l'exception toutefois du géo-référencement basé sur certains *MNE* dont la précision est insuffisante dans notre contexte de reliefs escarpés.

Nous traiterons ici uniquement des algorithmes de classification.

1. La classification

La classification a pour but d'attribuer une classe (ou label) choisie au sein d'un jeu de classes (par nature limité) à chaque élément d'un jeu de données. En ce qui concerne le traitement d'images, nous l'avons évoqué dans les chapitres précédents, la classification tente d'attribuer une classe à chaque pixel de l'image.

Mais la classification désigne tout à la fois cette opération de classification et l'image synthétique résultante dont la valeur de chaque pixel représente une classe (la classification finale).

Il existe deux types d'algorithmes de classification, les algorithmes non-supervisés et les algorithmes supervisés.

Les premiers cherchent à regrouper automatiquement les pixels d'une image dans des classes, dont le nombre fini peut être fixé ou non au départ. Cette opération est basée sur l'analyse des propriétés spectrales des pixels, qu'elle soit statistique ou géométrique. Ces méthodes sont rapides mais leur inconvénient provient du fait que les classes ne sont pas définies *a priori*, l'étiquetage des classes se faisant *a posteriori*. Ce point peut être un avantage car il permet de trouver les principales classes, en revanche certaines classes moins discriminantes mais d'importance sémantique pourraient être oubliées. Qui plus est, l'étiquetage des classes est une opération délicate car celles-ci peuvent être d'origine diverses et la cartographie finale n'avoir que peu de sens.

Les algorithmes supervisés sont, quant à eux, basés sur un jeu de données d'apprentissage, issu d'observations sur le terrain ou par photo-interprétation, pour lequel chaque pixel s'est vu

attribuer une classe au sein d'un jeu de classes prédéfini. Pour tous les pixels de l'image, l'algorithme cherche à attribuer la même classe que les pixels les plus proches de la base d'apprentissage. Cette méthode est sensible à la constitution du jeu de classes qui doit être représentatif des différentes classes effectivement présente mais présente l'avantage de produire des cartographies intelligibles dont la sémantique est maîtrisée dès le départ.

L'élément clé de la plupart de ces algorithmes est l'apprentissage (machine learning).

Plusieurs approches ont été mises au point pour résoudre le problème de la prise de décision de l'appartenance d'un pixel à une classe, celles-ci sont basées essentiellement sur des hypothèses probabilistes (*classificateur naïf de Bayes, maximum de vraisemblance, ...*), sur la notion de proximité (*K plus proches voisins, distance minimum, mahalanobis, parallélépipède*), sur une recherche de structure (*arbre de décision, réseaux de neurones, ...*), ou encore sur la recherche de frontières entre les classes - approche par discrimination (*SVM*).

2. Mesure de la qualité

Pour valider le résultat d'une opération de classification, une mesure *a posteriori* de la qualité de la classification finale est effectuée. Cette mesure est qualifiée de précision (accuracy), mais ce terme peut prêter à confusion puisque ce terme est parfois utilisé pour qualifier la résolution spatiale d'une image. Nous parlons ici de la mesure de la qualité de la classification finale qui consiste à vérifier dans quelle mesure les pixels de l'image sont bien classifiés.

La mesure de la qualité des classifications est une étape incontournable qui doit être systématique. En effet, le succès d'une classification n'est jamais garanti tant il est fortement lié à la qualité de chaque maillon de la chaîne de traitements.

Ce succès n'est pas lié uniquement aux performances de l'algorithme de classification, mais aussi aux données de vérité terrain, à l'adéquation des propriétés des images avec les hypothèses théoriques des algorithmes, aux réglages des paramètres, etc.

Cette mesure est basée sur le calcul *a posteriori* de la matrice de confusion. Celle-ci est un tableau à deux entrées avec en abscisse les classes de la vérité terrain et en ordonnée les classes issues de l'image de classification.

Tous les pixels de la base de contrôle (issue de la vérité terrain) sont comptabilisés à l'intersection de la colonne correspondant à la classe spécifiée par la vérité terrain et de la ligne correspondant à la classe qui leur a été attribuée par l'algorithme de classification.

Les pixels bien classifiés sont donc comptabilisés dans la diagonale principale.

Cl.	Vérité terrain					Total
	UZ	MF	HF	PF	LV	
UZ	372 [58,03%]	31 [4,14%]	81 [11,17%]	21 [2,82%]	1 [0,14%]	506 [14,22%]
MF	138 [21,53%]	716 [95,72%]	21 [2,90%]	41 [5,51%]	0 [0,00%]	916 [25,74%]
HF	34 [5,30%]	0 [0,00%]	613 [84,55%]	0 [0,00%]	0 [0,00%]	647 [18,18%]
PF	1 [0,16%]	1 [0,13%]	0 [0,00%]	663 [89,11%]	71 [10,14%]	736 [20,69%]
LV	21 [3,28%]	0 [0,00%]	0 [0,00%]	19 [2,55%]	628 [89,71%]	668 [18,77%]
Non-cl.	75 [11,70%]	0 [0,00%]	10 [1,38%]	0 [0,00%]	0 [0,00%]	85 [2,39%]
Total	641 [100,00%]	748 [100,00%]	725 [100,00%]	744 [100,00%]	700 [100,00%]	3558 [100,00%]

tableau 23 Exemple de matrice de confusion : Pixel [%]

Le *tableau 23* est un exemple de matrice de confusion. On peut voir que la classe *MF* est la mieux classifiée, 716 pixels, soient 95.72% des pixels de cette classe dans la base de contrôle ont été correctement classifiés.

La colonne *MF* qui est la distribution de tous les pixels de la classe *MF* de la base de contrôle nous informe que 4.14% d'entre eux ont été incorrectement attribués à la classe *UZ*, et 0.13% à la classe *PF*.

La ligne *MF* qui synthétise les classes effectives de tous les pixels de la base de contrôle ayant été classifiés dans la classe *MF* nous informe que 21.53% d'entre eux sont en réalité de la classe *UZ*, 2.9% sont de la classe *HF* et 5.51% de la classe *PF*.

La matrice de confusion permet donc de synthétiser les succès de classification (diagonale principale) par classe mais aussi les confusions interclasses (éléments hors diagonale).

Cependant, il n'est pas aisé de comparer des matrices de confusion, il faut donc utiliser des mesures de la précision. Une première étape consiste à calculer les erreurs d'omission et les précisions producteur de chaque classe de la base de contrôle (*tableau 24*) ainsi que les erreurs de commission et la précision utilisateur de chaque classe attribuée par l'algorithme de classification (*tableau 25*).

	UZ	MF	HF	PF	LV
Om.	269/641 [41,97%]	32/748 [4,28%]	112/725 [15,45%]	81/744 [10,89%]	72/700 [10,29%]
Prod. Acc.	372/641 [58,03%]	716/748 [95,72%]	613/725 [84,55%]	663/744 [89,11%]	628/700 [89,71%]

tableau 24 Erreurs d'omission et précision « producteur »

L'erreur d'omission est le rapport du nombre de pixels mal classifiés d'une classe de la base de contrôle par le nombre total de pixels de cette classe.

La précision « producteur » est au contraire le rapport du nombre de pixels bien classifiés d'une classe de la base de contrôle par le nombre total de pixels de cette classe.

	Com.		User Acc.	
UZ	134/506 [26,48%]	372/506 [73,52%]		
MF	200/916 [21,83%]	716/916 [78,17%]		
HF	34/647 [5,26%]	613/647 [94,74%]		
PF	73/736 [9,92%]	663/736 [90,08%]		
LV	40/668 [5,99%]	628/668 [94,01%]		

tableau 25 Erreurs de commission et précision « utilisateur »

L'erreur de commission est le rapport du nombre de pixels classifiés dans une classe mais qui n'appartiennent pas réellement à cette classe par le nombre de pixels de cette classe.

La précision « utilisateur » est le rapport entre le nombre de pixels d'une classe qui appartiennent effectivement à cette classe par le nombre de pixels de cette classe.

En ce qui concerne les mesures globales de qualité des classifications, les plus couramment utilisées et reconnues sont le Mean Producer Accuracy (MPA), la précision globale ou Overall Accuracy (OA) et le coefficient Kappa.

Le MPA correspond à la moyenne des précisions « producteur » de chaque classe, le OA est le rapport du nombre de pixels bien classifiés (ceux de la diagonale principale) par le nombre total de pixels. Le coefficient Kappa est un indice plus évolué qui prend en compte le fait que si tous les pixels de la base de contrôle ont été classifiés de façon aléatoire, la mesure de la précision ne sera pas nulle.

Notre matrice de confusion obtient un MPA de 83.42% et un OA de 2992/3558pixels soit 84.09% et un coefficient Kappa de 0.8019.

Ces outils de mesure sont couramment utilisés en télédétection mais il faut y mettre un bémol.

Si l'on se réfère à la matrice de confusion, une classification parfaite est une classification qui aurait 100% des pixels de la base de contrôle correctement classifiés (tous les éléments de la

diagonale principale sont égaux à 100%). Dans ce cas, les indices précédents seront maximums.

Or, cela n'implique pas forcément que l'intégralité des pixels de l'image soient bien classifiés. Pour que la matrice de confusion soit représentative de la qualité de la classification, la base de contrôle doit être un échantillon représentatif de l'ensemble des pixels de l'image, ce qui est toujours délicat à obtenir.

Il faut donc prendre un soin particulier à la constitution de la base de contrôle afin que les mesures de précisions soient les plus réalistes possibles.

Cela implique aussi que les valeurs de précisions tirées de la matrice de confusion soient maniées avec précaution. Un biais est quasiment systématique et il est toujours difficile d'évaluer la pertinence d'une mesure de précision sans connaître au moins le protocole de vérité terrain et la constitution des bases de contrôle.

3. Comparaison multi-échelle

La comparaison, au sens strict du terme, des algorithmes de classification n'a pas beaucoup de sens, ce sont des outils et cela reviendrait à comparer un pinceau avec un autre pinceau. Comment affirmer que l'un est meilleur que l'autre ? Cela dépend évidemment du résultat que l'on veut obtenir et de la maîtrise que l'on en a. Tout au plus est-il possible d'affirmer que dans certains cas l'un est plus adapté que l'autre.

Il existe deux cas de figure où l'on effectue une comparaison d'algorithmes de classification : lorsque l'on cherche à démontrer quel est l'outil le plus adapté à une problématique donnée avec un jeu de données particulier, et lorsque l'on essaie de démontrer qu'un algorithme est plus performant que d'autres.

Le premier cas est effectué à chaque fois que je suis confronté à une configuration nouvelle, me permettant de vérifier et valider l'utilisation d'un algorithme parmi tous. Dans ce cas, la comparaison se fait avec des conditions identiques, sur un même milieu, un même jeu de classes et celui obtenant la meilleure qualité de classification est validé.

Le second cas est plus complexe à mettre en œuvre si l'on essaie d'être exhaustif, une comparaison avec des algorithmes unanimement reconnus et utilisés est souvent suffisante.

En 2003 une première étude de comparaison d'algorithmes de classifications sur le milieu corallien a été effectuée (Capolsini *et al.* 2003). Des images SPOT-HRV, Landsat 7/ETM+, ASTER, Ikonos et MASTER (MODIS-ASTER simulator) ont été classifiés par les algorithmes *maximum de vraisemblance*, *distance de Mahalanobis*, *réseaux de neurones*, et *logique floue* selon un jeu de classes de 6 niveaux de complexité de l'habitat corallien.

Une seconde étude de comparaison d'algorithmes a été effectuée sur des écosystèmes différents et des échelles différentes afin d'avoir un point de vue général et complet des algorithmes testés sur les milieux naturels polynésiens en général.

Cette étude compare l'algorithme *SVM* avec les autres algorithmes de classification récents que sont les algorithmes *naïve bayes*, l'algorithme *C4.5*, l'algorithme *random forest*, l'algorithme *boosted regression tree* et l'algorithme des *K plus proches voisins*.

Le site d'étude est l'île de Moorea où trois types d'écosystèmes ont été étudiés : les écosystèmes anthropogéniques terrestres (*ATE*), les écosystèmes naturels terrestres (*NTE*) que sont les forêts primaires et les écosystèmes marins (*ME*).

L'intégralité des données satellites existantes et disponibles sur l'île de Moorea a été utilisée. Le descriptif de ces images est résumé dans le *tableau 26* suivant. Les dates vont de 1986 à

2010, les résolutions de 30 mètres à 50 cm et le nombre de bandes de 2 à 8. Des données optiques ainsi que radar ont été utilisées.

Capteur	Résolution (m)	Date	nbre bandes
Landsat-7 ETM+	30.0	2000	3
SPOT	20.0	1986	3
AirSAR	5.00	2000	4
TerraSAR-X	2.75	2009	2
Quickbird	0.60	2006	4
WorldView-2	0.50	2010	8

tableau 26 Caractéristiques des capteurs

Le nombre de classes est fixé à 4 dans tous les cas de figure car des études ont montré qu'il a une influence directe sur la précision des classifications. Un jeu de classes de communautés a été défini pour les images décimétriques comme Landsat et Spot, celui-ci tend vers un jeu de classes d'espèces pour les images métriques et sub-métriques.

De même, le nombre de pixels d'entraînement et de vérification est fixé à 250 pixels pour chacune des 4 classes afin de s'affranchir de l'influence de ce paramètre.

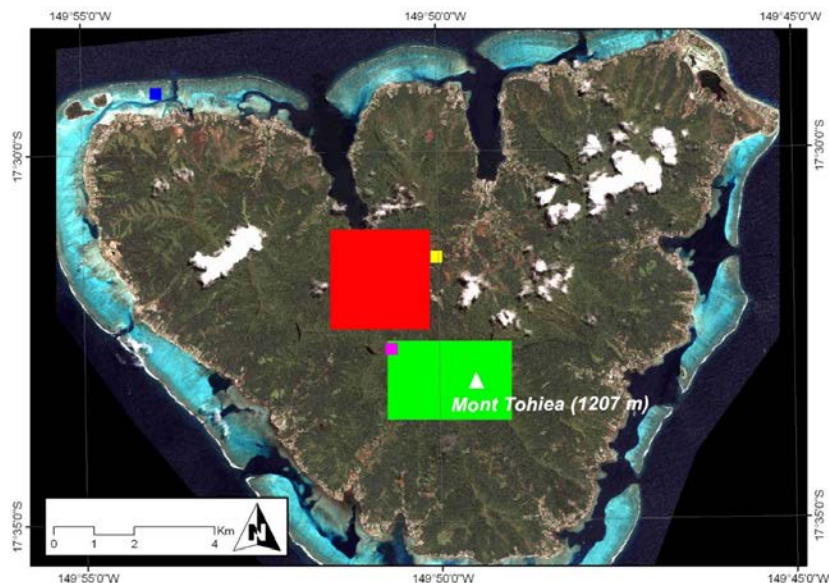


figure 31 Zones d'études - Moorea

La taille des zones d'étude a été choisie en fonction de la résolution des images afin de travailler sur un nombre de pixels fixe (figure 31) : toute l'île de Moorea pour les images décimétriques, et des zones plus petites pour les images métriques (carré rouge et rectangle vert) et sub-métriques (carrés bleus, jaune et violet)

Aucune différence significative n'a été trouvée sur la précision globale des classifications entre les différents types d'écosystème considérés (*ATE*, *NTE* et *ME*) ce qui tend à montrer que les algorithmes de classification supervisés développés ces dernières années sont à même de manipuler des données complexes de façon équivalente et efficace (figure 32, figure 33, figure 34).

L'algorithme *SVM* montre ici qu'il est un algorithme efficace pour classifier des images d'écosystèmes complexes puisqu'il obtient de meilleures performances dans 75% des expériences. L'algorithme *Random Forest* obtient les performances les moins satisfaisantes car il n'arrive en tête que dans un seul cas.

De façon générale, une réduction du nombre de bandes ou un accroissement du niveau de détails, entraîne une dégradation de la précision des classifications.

Cependant, la précision obtenue par l'algorithme *SVM* décroît dans une moindre mesure par rapport aux autres algorithmes lorsque le niveau de complexité augmente.

Ces résultats font de l'algorithme *SVM* un choix particulièrement pertinent pour les études sur les milieux naturels terrestres des îles polynésiennes, par nature fortement hétérogènes et complexes.

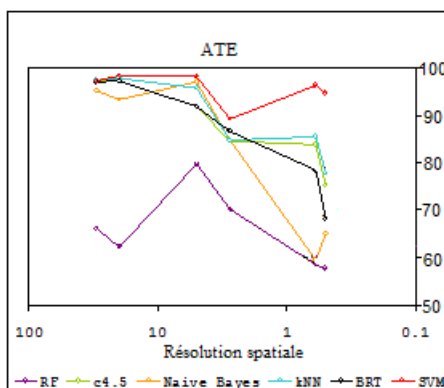


figure 32 Précision globale (OA) comparative sur les ATE

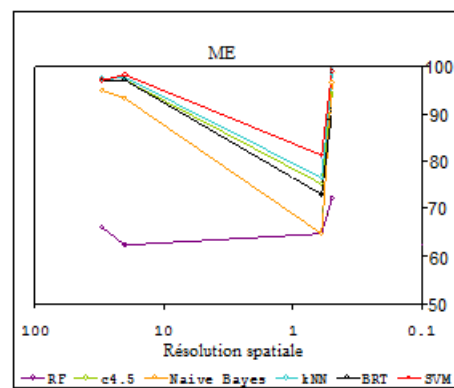


figure 33 Précision globale (OA) comparative sur les ME

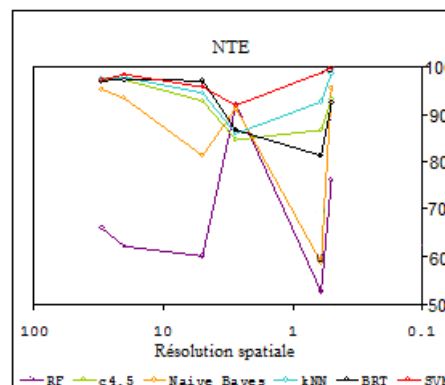


figure 34 Précision globale (OA) comparative sur les NTE

Références : (Pouteau, Collin et al. isrse2011).

B. L'ALGORITHME SVM

La plupart des comparaisons entre les algorithmes de classification classiques et l'algorithme *SVM* montrent que ce dernier l'emporte dans la majorité des cas de figure. La pertinence de l'utilisation de cet algorithme sur notre thématique a été validée par l'analyse comparative multi-échelle précédente. Ce chapitre présente une analyse fonctionnelle plus poussée pour en comprendre les différents aspects et subtilités.

1. Concepts

L'algorithme *SVM*, a été introduit à l'origine comme un algorithme de classification binaire. Il repose sur l'optimisation d'un hyperplan de séparation entre les échantillons de deux classes (présence/absence) dans un espace multidimensionnel.

Pour ce faire, il cherche à maximiser la distance entre les échantillons les plus proches (les vecteurs supports) et l'hyperplan de séparation.

Lorsque les échantillons ne sont pas linéairement séparables, les vecteurs sont projetés dans un espace de dimension supérieure dans lequel ils sont linéairement séparables (*figure 35*). Pour cela, une fonction noyaux (*Kernel*) est utilisée, puis l'hyperplan de séparation est déterminé grâce à une fonction d'optimisation.

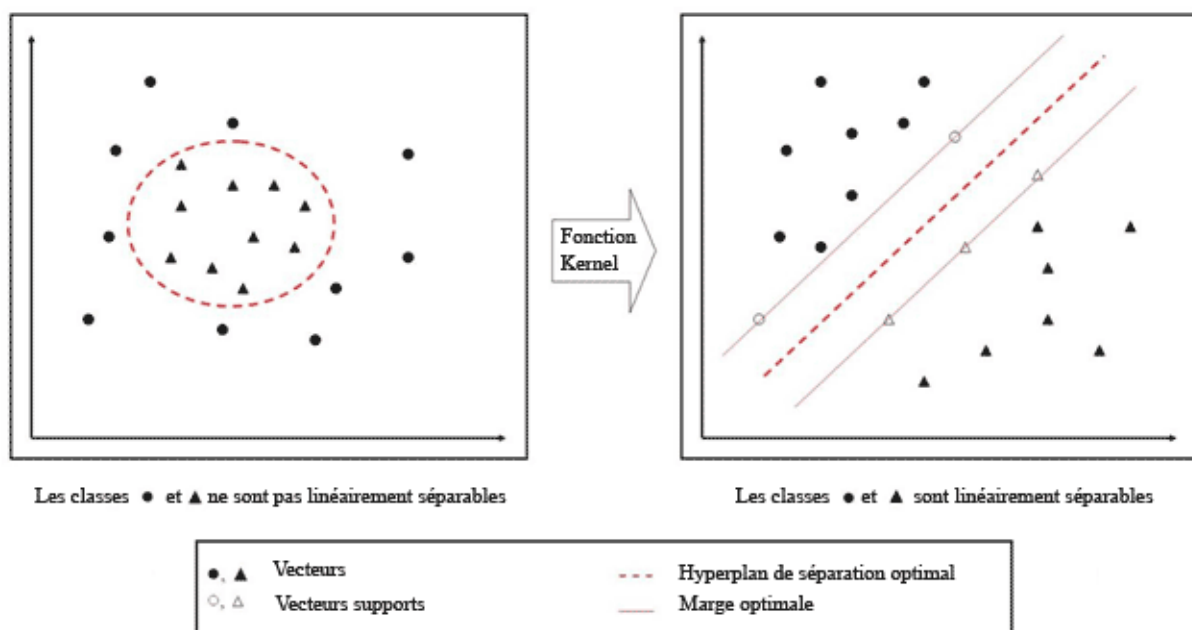


figure 35 Concept de l'algorithme SVM.

Plusieurs fonctions noyaux ont été proposées dans la littérature, citons les noyaux *RBF* (Gaussian Radial Basis), Linéaire, Sigmoid et Polynomial. Ceux-ci ont été testés et comparés dans de nombreuses études, le noyau *RBF* est choisi dans la plupart des cas, car il présente l'avantage de n'avoir qu'un seul paramètre et d'obtenir malgré tout de bons résultats.

L'extension multi-classe de l'algorithme *SVM* binaire est classiquement orientée selon deux stratégies :

- La méthode « **un contre tous** » qui cherche à optimiser Q hyperplans séparant chaque classe des $Q - 1$ autres classes
- La méthode « **un contre un** » qui cherche à optimiser les $\frac{Q(Q-1)}{2}$ hyperplans séparant chaque couple de classes.

La classe finale est obtenue par la technique de vote majoritaire.

La stratégie la plus couramment utilisée dans la littérature scientifique est la méthode « un contre un ».

Le principal inconvénient de l'algorithme *SVM* est qu'il requiert un temps de calcul conséquent (programmation quadratique) et que celui-ci augmente de façon logarithmique avec le nombre d'échantillons d'apprentissages.

Le principal avantage des algorithmes *SVM* est qu'ils n'estiment pas la distribution des données, permettant ainsi de classifier des données hétérogènes de caractéristiques statistiques totalement différentes. Ils sont aussi particulièrement adaptés pour traiter des cas non-linéairement séparables ou de grande dimension.

Par son principe même, l'algorithme *SVM* est insensible aux échantillons peu pertinents (éloignés de la frontière entre 2 classes) car l'hyperplan optimal n'est basé que sur quelques pixels. Parmi ceux-ci, les échantillons incohérents sont écartés par un paramètre de relaxation, rendant la méthode plus robuste au bruit.

Peu d'échantillons d'apprentissage sont donc en théorie nécessaires pour optimiser l'hyperplan de séparation des classes, pourvu qu'ils soient représentatifs de la frontière entre les classes. C'est une propriété intéressante lorsque l'on ne dispose que de peu d'échantillons terrains, et cela permet de surcroît d'obtenir des temps de calcul raisonnables.

Nous verrons que toutes les caractéristiques de cet algorithme lui confèrent un grand intérêt pour la suite des travaux présentés dans ce document.

2. Optimisation des paramètres

a. Par validation croisée

La première étape de l'utilisation de l'algorithme *SVM* est l'optimisation des paramètres *SVM*. Ceux-ci dépendent essentiellement du type de noyau choisi, chaque noyau ayant ses paramètres propres.

La méthode la plus utilisée pour déterminer les paramètres *SVM* optimaux est la technique de validation croisée.

Celle-ci consiste à diviser les pixels d'apprentissage en n sous-échantillons. Puis le *SVM* est entraîné sur $n - 1$ sous-échantillons, la précision est ensuite mesurée sur l'échantillon restant, cela permet d'éviter les problèmes de sur-apprentissage.

Cette technique est répétée pour chacun des paramètres selon une grille de recherche. Cette grille de recherche sera exponentielle dans un premier temps, puis linéaire pour affiner les résultats.

b. Par auto-configuration itérative.

La technique précédente obtenant d'excellents résultats, elle a été retenue dans la plupart des cas. Elle a néanmoins le défaut d'être très longue. En effet, pour trouver les valeurs optimales de deux paramètres *SVM*, et en mettant en œuvre une grille avec 5 valeurs par paramètres selon une grille grossière et une grille fine, il faut effectuer 50 classifications *SVM* pour choisir le jeu de paramètres optimal.

Des techniques d'optimisations *SVM* moins chronophages font à l'heure actuelle l'objet de nombreuses études. D'ici là, on peut se demander dans quels cas il est nécessaire de ré-optimiser les paramètres *SVM*. Cette question a fait l'objet d'une étude afin de mieux

comprendre la relation entre les données d'apprentissage et la nécessité d'optimisation des paramètres dans la cadre de la classification de textures.

La librairie Orphéo Toolbox (*OTB*) développée par le CNES a été utilisée. Cette librairie gratuite propose une implémentation des algorithmes *SVM* ainsi qu'une procédure de calibration automatique des paramètres *SVM*.

La méthode itérative d'optimisation des paramètres de la librairie *OTB* nécessite deux paramètres : le nombre maximum d'itérations ainsi que le type de noyau Kernel à optimiser.

Cherchons dans un premier temps à savoir quelle est l'influence du nombre d'itérations sur les résultats de classification de vignettes de texture.

Une base d'apprentissage de 60 vignettes de textures de 32x32 pixels (10 vignettes pour chacune des 6 classes) est utilisée pour la phase d'apprentissage de l'algorithme *SVM* conduisant au paramétrage de l'hyperplan de séparation.

Cette même base de vignette de texture est utilisée pour la phase de reconnaissance de l'algorithme *SVM*. Le taux de vignettes classifiées dans la bonne classe nous donne une mesure de l'efficacité de la classification, c'est le *GRP* (good recognition percentage).

Cette procédure est effectuée pour un nombre d'itérations allant de 10 à 70, pour chaque noyau de l'algorithme *SVM*.

La *figure 36* montre le taux de bonnes reconnaissances (*GRP*) obtenu pour les 4 principaux noyaux. Le *tableau 27* présente les temps de calcul nécessaires pour l'auto-configuration des paramètres et l'apprentissage en fonction du noyau utilisé.

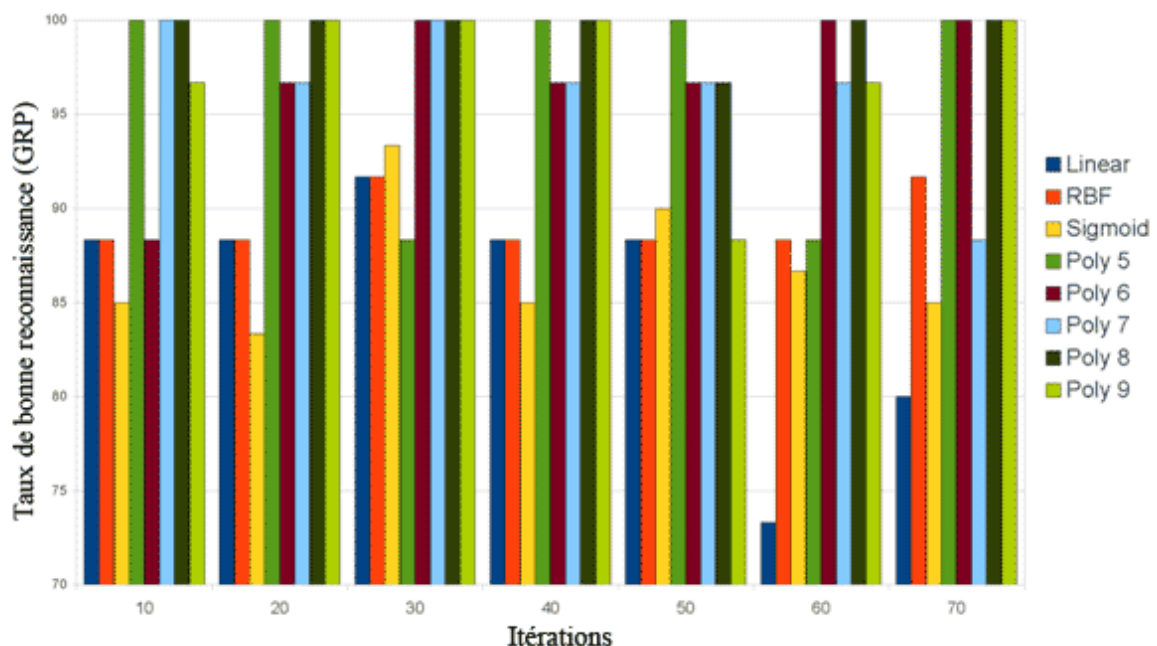


figure 36 Taux de bonne reconnaissance en fonction du nombre d'itérations

Cette expérience montre que le nombre d'itération de la méthode automatique de configuration des paramètres *SVM* donne des résultats relativement similaires, que les noyaux polynomiaux sont les plus efficaces, mais qu'ils nécessitent le temps de calcul le plus long.

Le noyau *RBF* s'avère être un bon compromis en termes de temps de calcul, de taux de bonnes reconnaissances et de complexité de mise en œuvre, avec un nombre d'itérations de 30 pour la configuration automatique des paramètres des *SVM*.

Kernel	5 itérations		70 itérations		30 itérations	
	60 attributs		8 attributs		60 attributs	
	60 vignettes		30 vignettes	1920 vignettes	30 vignettes	1920 vignettes
Linear	1"	21"	1"	27"	1"	45"
RBF	3"	1' 36"	1"	13' 9"	7"	28'
Sigmoid	40"	3h 12'	49"	12h	7'	28h 48'
Polynomial	1'	5h 44'	10' 48"	20h 40'	8'	31h 45'

tableau 27 Temps de calcul pour l'auto-configuration et l'apprentissage SVM.

Nous l'avons vu, les temps de calcul pour l'optimisation des paramètres de l'algorithm SVM et pour l'apprentissage sont importants et ceux-ci augmentent rapidement avec le nombre de vignettes d'apprentissage.

Est-il nécessaire d'effectuer systématiquement ces deux opérations, ou peut-on réutiliser un jeu de paramètres ? Sous quelles conditions ?

Pour répondre à cette question, le noyau *RBF* et une optimisation avec 30 itérations sont utilisés. L'expérience est basée sur une base de données de 3840 vignettes de texture (640 vignettes pour chacune des 6 classes) sur lesquelles on calcule les 15 attributs de textures d'Haralick et Shapiro¹⁶ pour chaque bande (soit 60 attributs de textures).

La moitié de cette base de données est utilisée pour optimiser les paramètres SVM et conduire l'apprentissage (1920 vignettes). L'autre moitié servira pour effectuer la reconnaissance.

Puis, l'expérience est réitérée en réduisant le nombre de vignettes d'apprentissage. La reconnaissance étant effectuée sur le reste des vignettes non encore utilisées.

Cette étape est conduite d'une part en ré-optimisant les paramètres SVM et d'autre part en réutilisant les mêmes paramètres que ceux trouvés à la première étape.

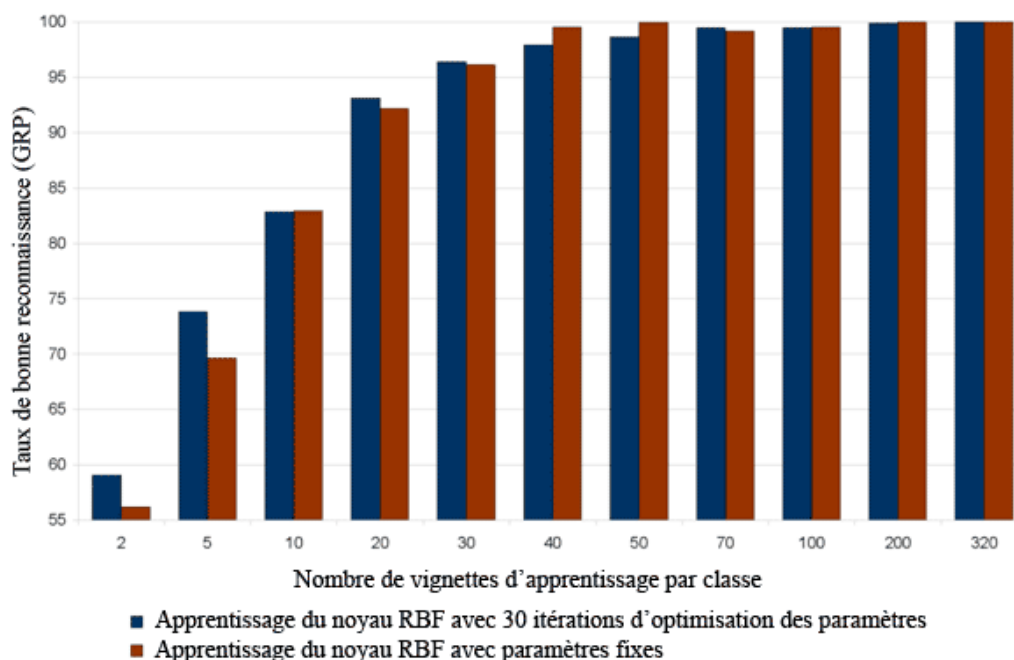


figure 37 Optimisation systématique vs optimisation unique

¹⁶R. H. Haralick and L. G. Shapiro, "Image segmentation techniques," Computer Vision Graphics Image Processing, vol. 29, pp. 100–132, 1985.

La *figure 37* compare les résultats obtenus lorsque l'on effectue une ré-optimisation systématique des paramètres avec ceux obtenus lorsque l'on utilise les paramètres optimisés au départ (optimisation unique).

Les résultats sont relativement similaires : les paramètres optimisés sur une base de données peuvent donc être réutilisés si l'on doit classifier un sous-ensemble de cette base de données.

Ce résultat peut-il s'appliquer si l'on réduit la base de données en diminuant le nombre d'attributs de textures ?

Les 15 attributs d'Haralick et Shapiro sont calculés pour chacune des quatre bandes des vignettes Quickbird (60 attributs de textures par vignette). Doit-on ré-optimiser les paramètres *SVM* si l'on ne désire utiliser qu'une partie des attributs de textures ?

L'optimisation est effectuée sur les 60 attributs de textures de la moitié des vignettes de la base de données (1920 vignettes). L'autre moitié servira pour effectuer la reconnaissance et vérifier les résultats.

Le nombre d'attributs de textures est ensuite réduit de 60 à 10 attributs, puis à 8. L'étape de reconnaissance est à chaque fois effectuée sur le restant des vignettes, d'une part en utilisant les paramètres trouvés lors de l'optimisation sur les 60 attributs, et d'autre part en ré-optimisant sur le nombre d'attributs réel.

La *figure 38* montre sans conteste qu'il est impératif de ré-optimiser les paramètres *SVM* à chaque fois que l'on modifie le nombre d'attributs de textures utilisés.

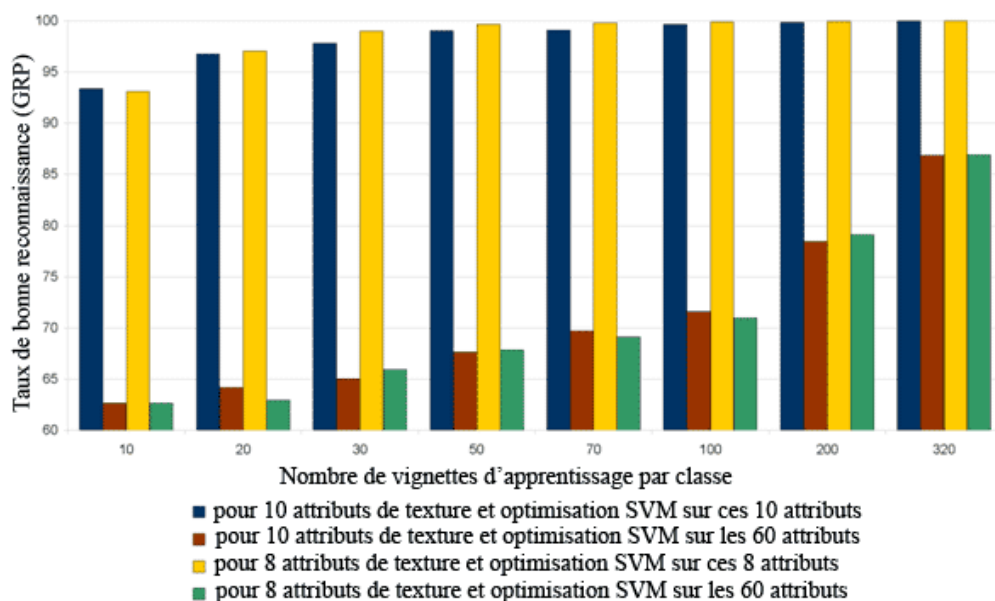


figure 38 Importance de la ré-optimisation des paramètres *SVM*

Références : (Chabrier et al. aprs2012).

3. Classification SVM de données radar

Nous l'avons vu, la donnée JPL-AirSAR en mode polarimétrique pour les bandes L et P , est livrée au format *MLC (Multi-look complex)* qui correspond à 9 vues. Pour chaque pixel, la matrice de cohérence est dérivée des paramètres de la matrice de Stokes.

Toute l'information polarimétrique est contenue dans cette matrice de cohérence, il existe différentes méthodes de classification de données radar polarimétrique : les plus classiques étant la technique par *maximum de vraisemblance* basée sur la distribution de *Wishart* ou sur la *décomposition H/A/α*.

Ces algorithmes ont toutefois montré des limites dans des cas complexes avec en particulier un grand nombre de bandes, or, comme nous l'avons vu, c'est lorsque la complexité du problème augmente que l'algorithme *SVM* fait la différence.

Nous cherchons donc à valider l'utilisation de l'algorithme *SVM* pour classifier des données radar par une comparaison entre celui-ci et l'algorithme *de Wishart*, spécifiquement développé pour la classification de données radar polarimétriques.

Ces deux algorithmes sont donc utilisés pour classifier les 9 éléments de la matrice de cohérence pour les bandes L et P ainsi que l'ensemble des 19 éléments de la matrice de cohérence pour le cas $L + P + C_{VV}$.

En ce qui concerne l'algorithme *SVM*, des tests préliminaires ont permis de sélectionner le noyau *RBF* et une stratégie « un contre un ».

La précision moyenne (*MPA*) des différentes classifications ainsi que les précisions pour chaque classe sont reportés dans le *tableau 28* ci-dessous :

Classes	L		P		L+P+C _{VV}	
	Wishart	SVM	Wishart	SVM	Wishart	SVM
Pinus	32%	74%	56%	71%	81%	99%
Falcata	69%	80%	37%	64%	75%	99%
Purau	48%	78%	61%	85%	75%	98%
Goyavier de chine	76%	88%	66%	74%	74%	100%
Lande à fougère	62%	93%	49%	88%	56%	98%
Marais	85%	97%	92%	93%	95%	99%
sol nu	90%	98%	91%	96%	91%	99%
MPA	66%	87%	62%	82%	78%	99%

tableau 28 Précisions par classe des classifications de *Wishart* et *SVM*

De façon générale, l'algorithme *SVM* donne des résultats supérieurs d'à peu près 20% par rapport à l'algorithme de *Wishart*. Cela peut s'expliquer par une différence entre la distribution de la donnée JPL-AirSAR et la distribution théorique de *Wishart*. Cette différence proviendrait des surfaces non homogènes avec présence de textures des images.

L'algorithme *SVM* présente donc un fort potentiel pour la classification de données radar polarimétriques même s'il n'a pas été initialement développé dans ce but. La *figure 39* présente la classification *SVM* finale de l'île de Tubuai utilisant les bandes C , L et P (le jeu de classes est celui du *tableau 28*).

Références : (Lardeux et al. igarss2006; Stoll et al. perf2006; Lardeux et al. tgrs2009)

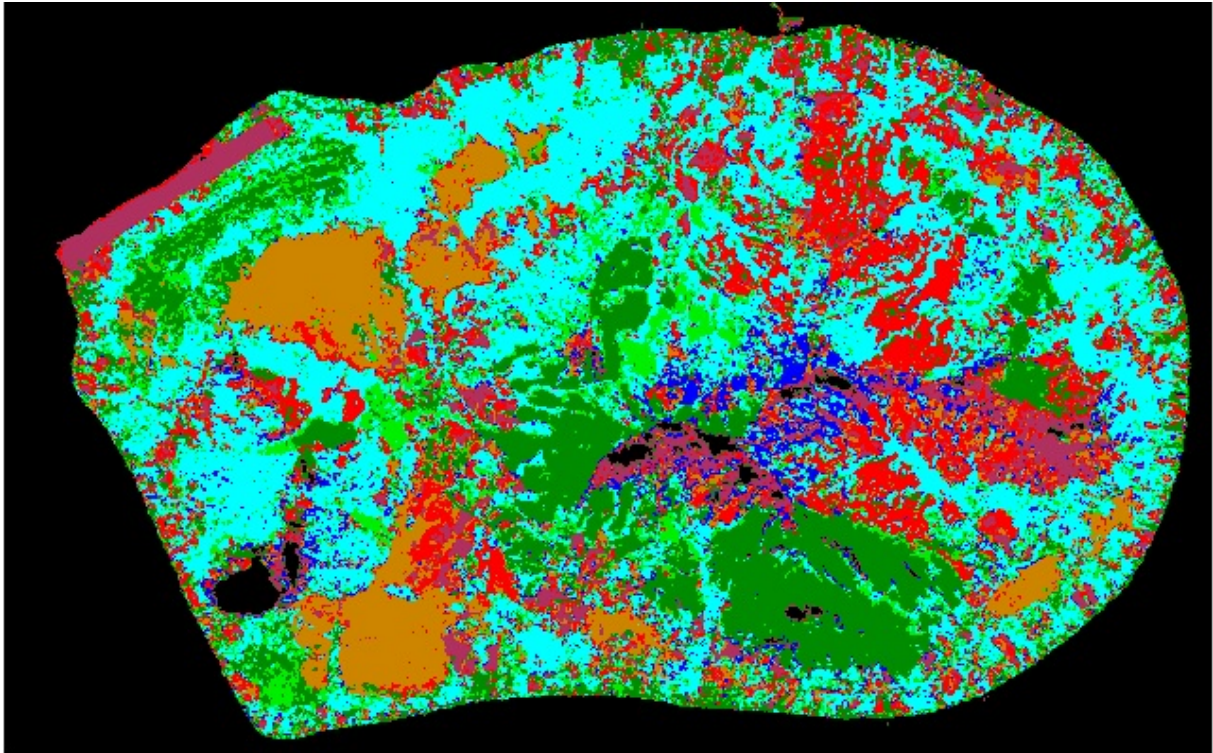


figure 39 Classification radar SVM - Tubuai

4. Vérité terrain

Les algorithmes de classification supervisés nécessitent une base de pixels de vérité terrain pour les phases d'apprentissage et de contrôle des résultats. Les pixels d'apprentissage servent à apporter à l'algorithme un échantillon de pixels représentatifs de chaque classe. Les pixels de contrôle permettent de contrôler la qualité de la classification par le calcul d'une matrice de confusion.

Une des méthodes usuelles d'échantillonnage terrain pour les algorithmes de classification supervisés consiste à repérer géographiquement des pixels purs (appartenant à une seule classe sans ambiguïté) ou des zones homogènes (dont tous les pixels appartiennent à une même classe) et ce pour chaque classe. Cependant, des études ont montré que l'échantillonnage de zones mixtes permet d'améliorer la précision des classifications en indiquant de façon plus précise la limite entre les classes.

L'algorithme *SVM* a un principe très différent des algorithmes probabilistes puisqu'il ne considère pas la distribution statistique des données. La méthodologie de vérité terrain a-t-elle une influence sur sa capacité de discrimination ?

Le principe de cet algorithme est de chercher l'hyperplan optimal séparant deux classes de pixels. Seuls les pixels d'apprentissage proches de celui-ci (les vecteurs supports) permettent d'affiner cet hyperplan de façon efficace, les pixels purs n'étant pas utiles.

Ainsi, contrairement aux algorithmes de classification probabilistes, il ne semble pas nécessaire d'avoir beaucoup de pixels d'apprentissage pour obtenir les meilleures performances, ce qui est d'autant plus appréciable que le temps de calcul augmente de façon logarithmique avec le nombre de pixels d'apprentissage.

Des études montrent pourtant que la précision de l'algorithme *SVM* s'améliore aussi avec l'augmentation du nombre de pixels d'apprentissage. Si ces études semblent à première vue contredire l'affirmation précédente, cela pourrait aussi s'expliquer par le fait que l'on trouve

ainsi statistiquement de plus en plus de vecteurs supports aidant à affiner l'hyperplan de séparation entre les classes.

Il conviendrait donc au contraire de trouver un nombre restreint de pixels qui soient fortement représentatifs de la frontière entre deux classes. Il est alors proposé de chercher sur le terrain des pixels représentatifs de ces vecteurs supports. L'hypothèse émise est que ces vecteurs supports sont présents à la frontière physique entre deux classes (entre deux types de végétation), sur ce que l'on appelle les écotones (figure 40).

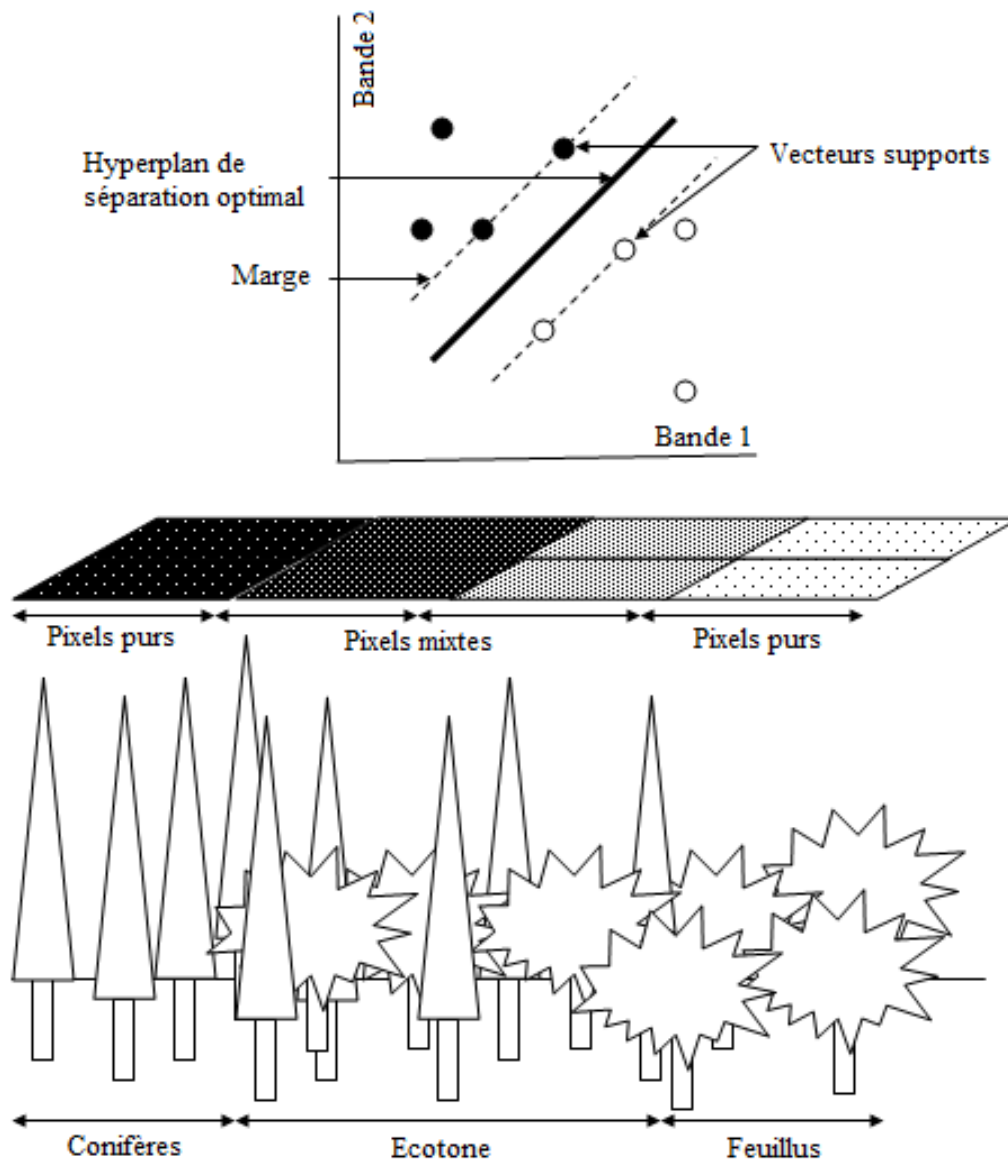


figure 40 Emplacement des écotones

Le site d'étude choisi est la planèze du mont Marau de l'île de Tahiti. La donnée disponible est une image Quickbird (4 bandes couleurs : rouge, vert, bleu, et infrarouge, résolution spatiale : 60cm). L'objectif est aussi de cartographier toute cette zone où l'on trouve une partie de la forêt de nuages de l'île de Tahiti, riche en espèces endémiques.

La première étape consiste à repérer ces zones d'intérêt sur nos images afin de pouvoir conduire efficacement la mission sur le terrain. Une analyse en composante principale (ACP)

est effectuée à cet effet, puis un algorithme de détection de contour de Sobel¹⁷ est appliqué sur la première composante de l'ACP.

Cette méthodologie permet de localiser les zones de transitions entre deux milieux. Deux emplacements de chaque côté de la frontière sont repérés, comme sur l'exemple suivant (figure 41) où l'on détecte la frontière entre les forêts de *Pinus Caribaea* (conifères) et *Falcata moluccana* (feuillus).

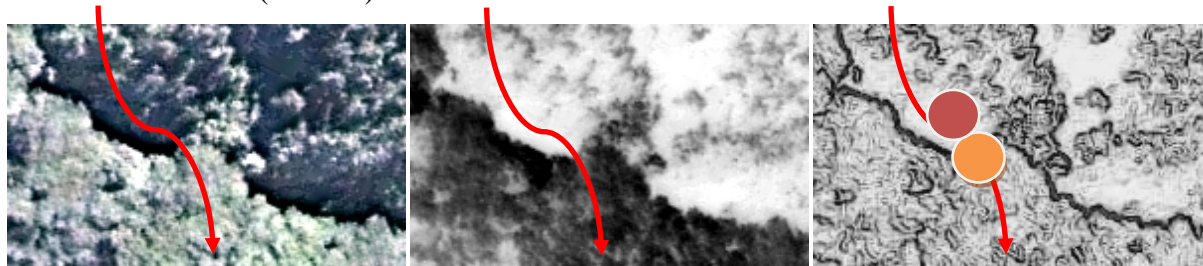


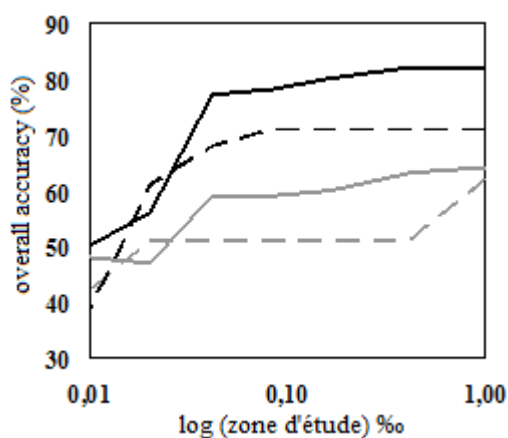
figure 41 Méthodologie de repérage et d'échantillonnage des écotones

Une mission terrain sur le domaine du mont Marau à Tahiti a permis de récolter des points sur les écotones ainsi que de part et d'autre de ceux-ci afin de constituer deux jeux de vérité terrain, l'un constitué de pixels sur les écotones, et l'autre de pixels purs.

Les algorithmes de *maximum de vraisemblance* et *SVM* ont tous deux été utilisés avec les deux jeux de donnée terrain (pixels purs et pixels sur écotones) en faisant varier le nombre de pixels d'apprentissage afin d'étudier l'influence de ce facteur.

La figure 42 compare les précisions des classifications obtenues en fonction du nombre de pixels d'apprentissage (%) dans les cas suivants :

- Classification *SVM* - Apprentissage sur les pixels purs (courbe en noir pointillée),
- Classification *SVM* - Apprentissage sur les écotones (courbe en noir continue),
- Classification par *maximum de vraisemblance* - Apprentissage sur les pixels purs (courbe en gris pointillée),
- Classification par *maximum de vraisemblance* - Apprentissage sur les écotones (courbe en gris continue).



- Max. de vraisemblance (gris)
- SVM (noir)
- Apprentissage sur pixels purs (pointillé)
- Apprentissage sur écotones (continu)

figure 42 Comparaison croisée algorithmes / échantillonnage terrain

¹⁷ I.E. Sobel, "Camera models and machine perception", PhD thesis, Stanford Univ., 1970.

La première constatation est que, l'augmentation du nombre de pixels d'apprentissage améliore systématiquement les résultats jusqu'à un plateau. De façon générale, il est donc nécessaire d'étudier le nombre optimum de pixels d'apprentissage permettant d'atteindre le plateau afin d'obtenir les meilleures classifications tout en limitant les échantillonnages terrain.

Ensuite, notons que dans les cas de l'algorithme de *maximum de vraisemblance* comme pour le SVM, le choix de pixels mixtes aux écotones aboutira à de bien meilleurs résultats, mais c'est particulièrement vrai pour l'algorithme SVM.

Cela corrobore l'hypothèse émise dans le cas de l'algorithme SVM, mais aussi les études sur l'algorithme de *maximum de vraisemblance* qui montrent que la prise de points de vérité terrain à la limite entre deux classes affine la frontière statistique entre celles-ci.

	Max. vraisemblance		SVM	
	Pixels purs	Pixels mixtes	Pixels purs	Pixels mixtes
OA (%)	62	71	64	82
Kappa	0.49	0.62	0.53	0.77

tableau 29 Max. de vraisemblance vs SVM / pixels purs vs pixels écotones

S'agissant du *tableau 29*, soulignons que le choix de la méthode de vérité terrain est un élément aussi déterminant sinon plus que le choix de la méthode de classification pour gagner en précision.

Le protocole de vérité terrain est donc primordial et influe directement sur la qualité des classifications. La prise de points d'apprentissage sur les écotones est particulièrement intéressante dans le cas de l'algorithme SVM.

La *figure 43* représente la classification du mont Marau à Tahiti obtenu par classification SVM et prise de points de vérité terrain sur les écotones.

Références : (Pouteau et al. igarss2010b).

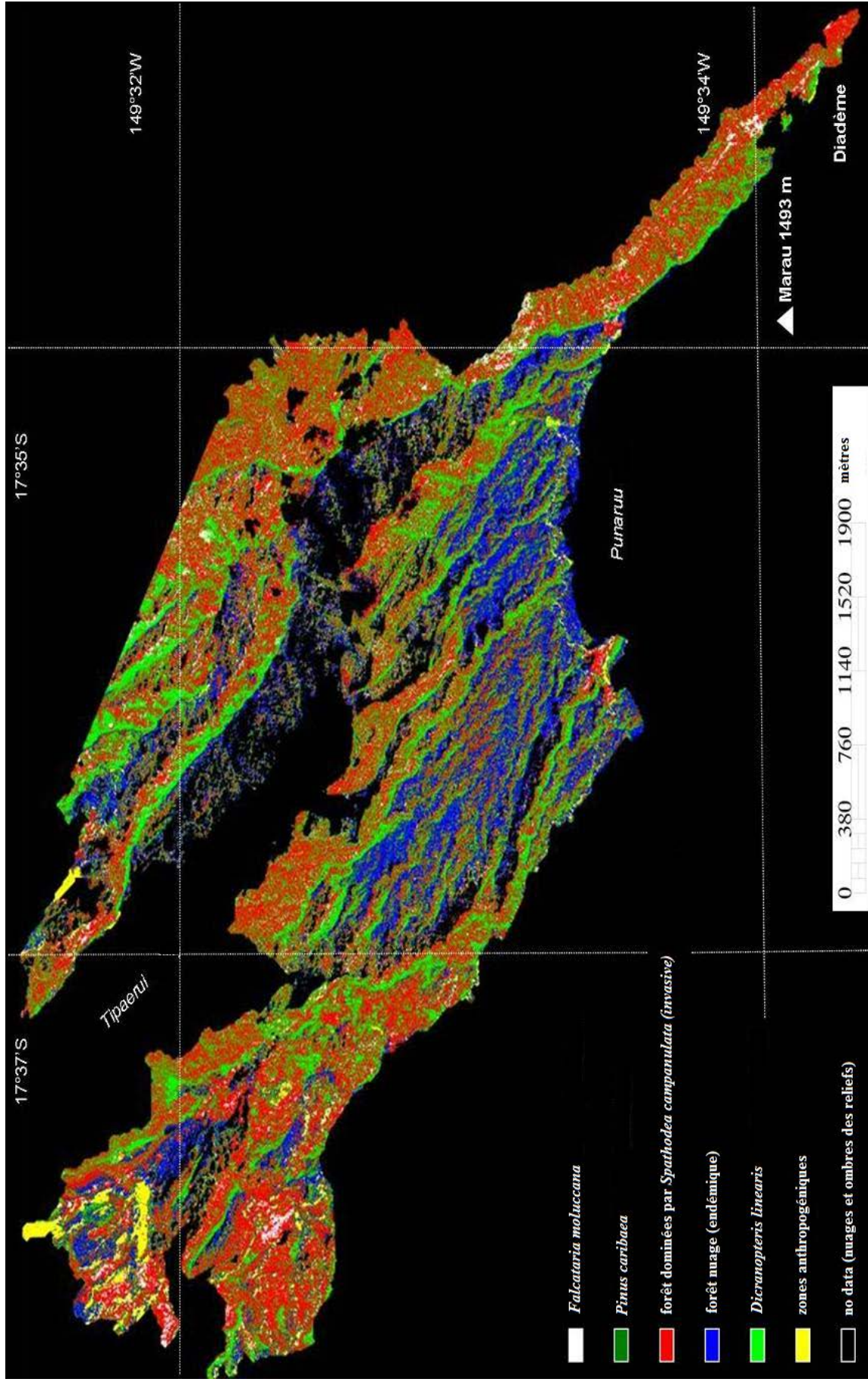


figure 43 Classification SVM avec échantillonnages sur les écotones– Mont Marau

IV. UTILISATION CONJOINTE DE DONNEES HETEROGENES

Comme établi précédemment, la donnée optique et la donnée radar présentent toutes deux un intérêt pour la classification du couvert végétal polynésien.

La donnée optique permet de discriminer un jeu de classe basé sur la réponse spectrale des différents types de végétation. Il est par exemple possible de discriminer les forêts d'arbres feuillus des conifères, ou les pâturages des landes à fougères.

La donnée radar permet quant à elle de discriminer les couverts végétaux selon des critères structurels comme la complexité du branchage et des troncs, la biomasse, le taux d'humidité. Ainsi il est possible par exemple de discriminer les pâturages, les champs d'ananas, les forêts composées d'arbres de grande taille, les forêts à hibiscus, etc.

Cependant, dans les deux cas, des confusions persistent entre certaines classes imposant l'utilisation d'un jeu de classes spécifique à chaque type d'images. Idéalement, le jeu de classes devrait être basé sur les différents types de paysages, de milieux naturels, ou les espèces végétales, en fonction de l'échelle considérée.

Or, la complexité du problème rend l'utilisation d'un seul type de donnée insuffisante pour classer correctement ce jeu de classes idéal. L'idée est alors d'utiliser les propriétés différentes mais complémentaires des données optiques et radars.

En généralisant ce raisonnement, le traitement simultané de plusieurs types de données est envisagé afin d'utiliser le maximum d'informations complémentaires sur la cible : des informations spectrales avec l'optique, des informations spatiales avec la texture, des informations structurelles avec le radar, mais aussi toute autre forme d'information additionnelle pertinente.

A. FUSION SVM MULTI-SOURCES

1. Intérêt

On l'aura compris, classer le milieu naturel polynésien dans toute sa complexité impose de multiplier les sources d'informations complémentaires comme les images optiques, radars, la texture, les descripteurs géomorphologiques, etc.

Avec son approche non-descriptive de la distribution des données, l'algorithme SVM est reconnu comme étant particulièrement adapté aux problèmes de classifications complexes comme lorsque l'on considère de nombreuses couches de données hétérogènes. Aussi tentante que soit l'idée de rajouter de plus en plus de couches d'informations différentes dans un algorithme de classification multi-sources SVM, cette approche atteint néanmoins rapidement ses limites lorsque des données redondantes ou non-pertinentes sont rajoutées.

Une approche différente nommée fusion multi-sources a fait l'objet de nombreuses études ces dernières années et celle qui a retenu notre attention a été proposée et testée avec succès par Waske et Benediktsson¹⁸.

Cette technique consiste dans un premier temps à effectuer des classifications mono-sources SVM sur chaque type d'image en conservant pour chacune la distance de chaque pixel à l'hyperplan. Puis dans un deuxième temps il s'agit de classer ces cartes de distances à l'hyperplan (*figure 44*).

¹⁸ B. Waske and J. A. Benediktsson, "Fusion of Support Vector Machines for Classification of Multisensor Data," IEEE Trans. Geosci. Remote Sens., vol. 45, pp. 3858-3866, 2007.

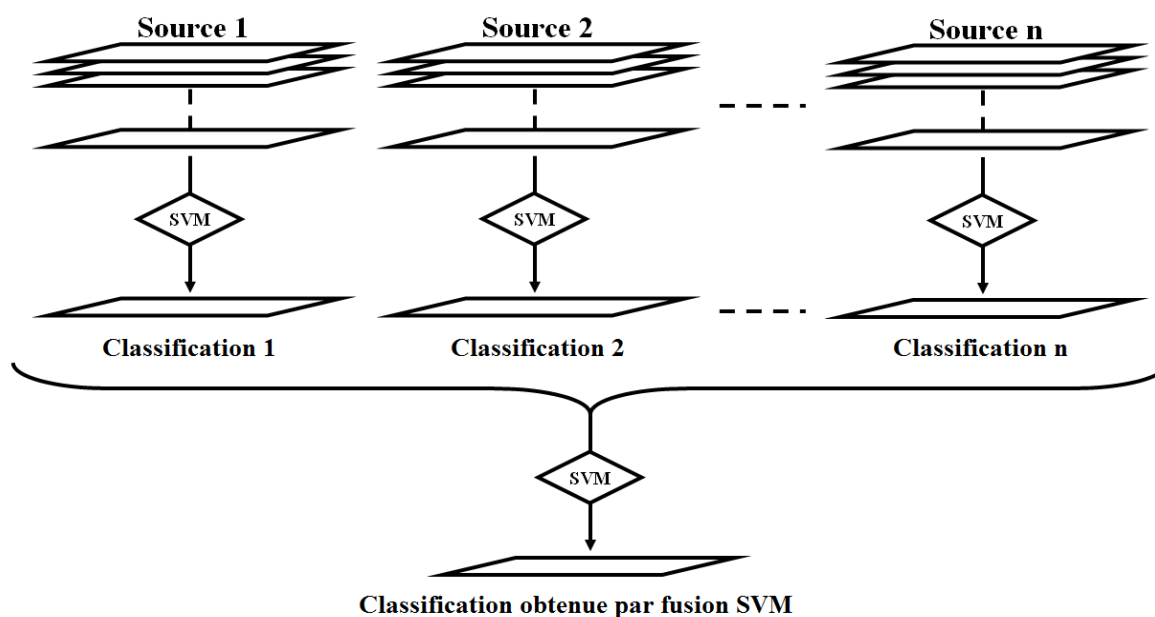


figure 44 Principe de la fusion SVM

Nous avons étudié la potentialité de ce schéma de fusion multi-sources sur l'île de Nuku-Hiva aux Marquises. Le jeu de données disponible est constitué d'une image optique, une image radar et un jeu d'indices géomorphologiques.

L'image Ikonos de 2005 a permis le calcul de quatre attributs de textures que sont la variance, le contraste, la dissimilarité, et le second moment angulaire. Trois tailles de fenêtre de 3x3, 9x9 et 15x15 ont été utilisées, correspondant respectivement à la micro-texture intra-arbres, la macro-texture intra-arbres et la texture inter-arbres.

Une étude de la corrélation des attributs de textures (matrice r^2) a été effectuée afin d'éliminer les redondances trop fortes.

Le deuxième type de données utilisées est la donnée JPL-AirSAR. Nous avons extrait de cette donnée les 10 indices polarimétriques les plus pertinents (cf. chapitre II.B.2). Un filtre de Frost a été appliqué pour réduire le chatoiement tout en préservant au mieux les détails.

Les indices géomorphologiques considérés sont l'élévation, la pente, l'exposition à l'Est (aux vents dominants) et l'indice CTI (cf. chapitre II.C.1).

Espèces	Strate	Statut
Acacia farnesiana (A)	S	Inv.
Casuarina equisetifolia subsp equisetifolia (Ca)	T	Ind.
Dicranopteris linearis (D)	H	Ind.
Falcataria moluccana (Fa)	T	Inv.
Ficus prolixa var prolixa (Fi)	T	Ind.
Hibiscus tiliaceus subsp tiliaceus (H)	T	Ind.
Inocarpus fagifer (I)	T	PI
Pandanus tectorius var tectorius (Pa)	T	Ind.
Psidium guajava (Ps)	S	Inv.
Sapindus saponaria (Sa)	T	Ind.
Schizostachyum glaucifolium (Sc)	T	PI

tableau 30 Jeu de classes Nuku-Hiva

Un jeu de classes d'espèces a été proposé (*tableau 30*). Pour chacune des 11 espèces la strate considérée (*T* : arbre, *S* : arbuste, *H* : herbacée), ainsi que le statut (*Inv* : envahissant, *PI* : introduction polynésienne, *Ind* : indigène) sont spécifiés.

La comparaison des 3 noyaux *SVM* les plus usités dans la littérature (*RBF*, *Polynomial* et *sigmoïde*) a permis de valider l'utilisation du noyau *RBF* qui obtient les meilleures précisions.

Le *tableau 31* illustre la précision globale des trois classifications mono-sources ainsi que celle de la fusion *SVM* multi-sources.

Précision	Ikonos (1m)	DEM	AirSAR	Ikonos (5m) + MNE + AirSAR
OA	54%	30%	20%	70%
Kappa	0.50	0.25	0.13	0.67

tableau 31 Précisions des classifications mono-sources et fusion *SVM* – Nuku-Hiva

La précision obtenue par la fusion *SVM* est bien meilleure que les précisions individuelles des classifications mono-sources. Cela corrobore les résultats de Waske and Benediktsson.

Analysons ces résultats classe par classe, la *figure 45* représente les précisions individuelles par classe de :

- La classification de l'image Ikonos additionnée de ses attributs de textures,
- la classification des indices géomorphologiques tirés du *MNE*,
- la classification de la donnée AirSAR,
- la fusion *SVM* des 3 classifications précédentes.

La donnée optique Ikonos avec ses attributs de textures est la plus adaptée pour discriminer un jeu de classes d'espèces (*OA* = 54%).

Le *MNE* additionné de ses indices géomorphologiques, malgré sa faible précision (*OA* = 30%), permet quant à lui de reconnaître certaines espèces écologiquement spécialisées comme *Casuarina equisetifolia* et *Dicranopteris linearis* présents exclusivement sur les affleurements rocheux et les crêtes (forte élévation, *CTI* faible), *Inocarpus fagifer* le long des rivières (*CTI* important) et *Sapindus saponaria* présent dans les forêts semi-xérophiles (*CTI* faible et sur forte pentes).

La donnée radar polarimétrique n'est pas adaptée à un jeu de classes d'espèces (*OA* = 20%), mais son absence du schéma de fusion *SVM* diminue les performances globales de façon significative, ce qui atteste de son importance.

Regardons de plus près deux cas particuliers de la *figure 45* :

Falcataria moluccana (flèche bleue) : cette espèce est mal classifiée sur les trois types de données séparément, en revanche elle obtient de très bons résultats (*OA* = 91%) lors de la fusion multi-sources *SVM*. Cela montre la capacité de cette dernière d'attribuer avec succès la bonne classe, alors que les résultats en mono-source ne sont pas bons. Le motif de confusion des classifications mono-sources est donc une information supplémentaire permettant de classer efficacement les pixels d'une classe en particulier. C'est là, à mon sens, le résultat le plus intéressant de la fusion *SVM* multi-sources.

Acacia farnesiana (flèche rouge) : cette espèce est bien discriminée sur la donnée optique, mais elle est mal discriminée sur la donnée AirSAR et sur les indices géomorphologiques tirés du *MNE*. La fusion multi-sources *SVM* dégrade ses performances par rapport à la classification mono-source de donnée optique. Pour cette classe en particulier, la fusion *SVM* n'apporte aucun bénéfice.

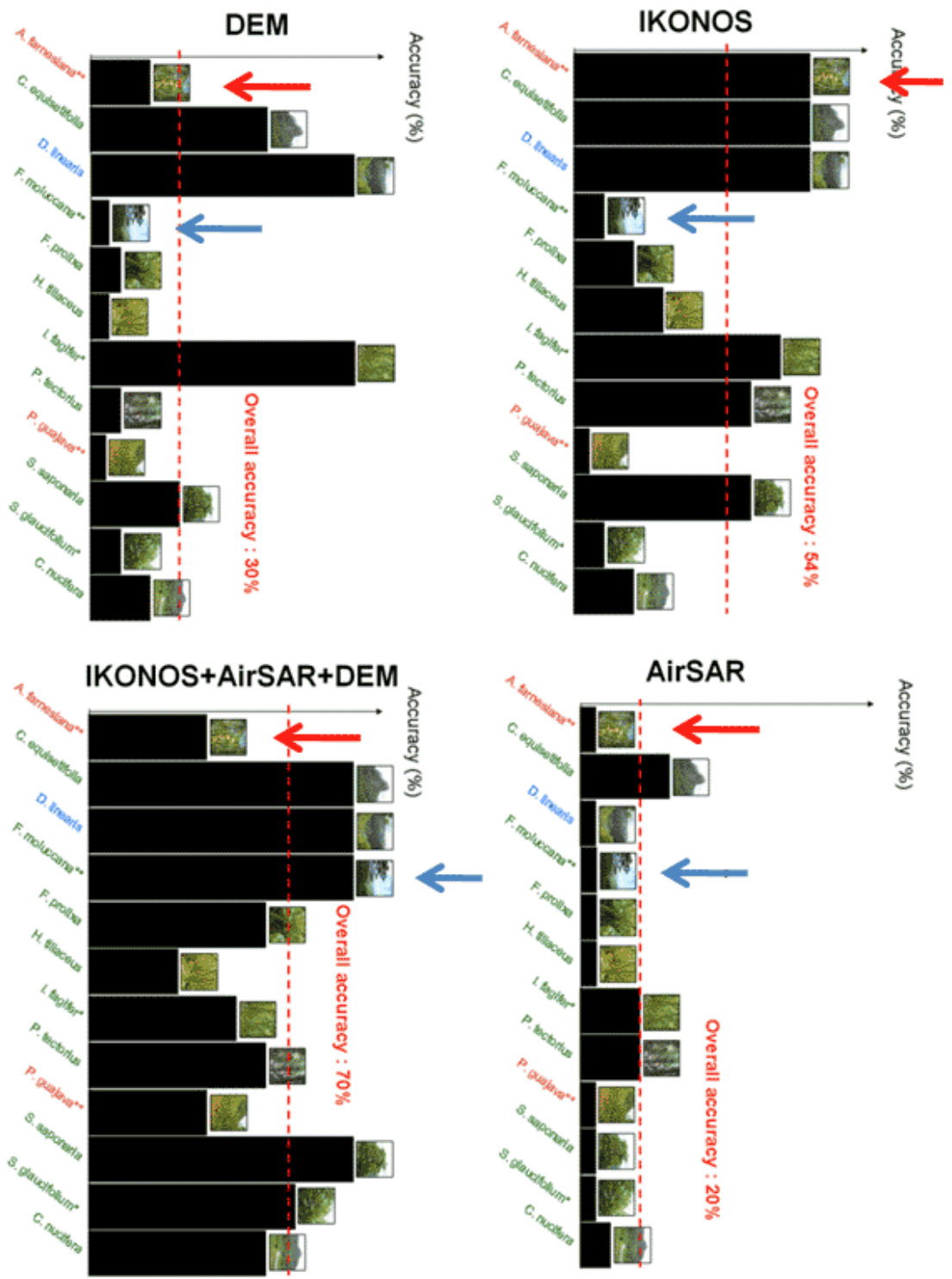


figure 45 Précision par classe des classifications mono-sources et fusion SVM

Pour valider ces résultats, nous avons appliqué cette méthodologie à l'île de Moorea. Le jeu de données consiste en une image SPOT, une image AirSAR et des indices topographiques (élévation, pente, et aspect) tirés du MNE de Moorea (figure 46).

Le tableau 32 suivant corrobore l'amélioration de la précision globale de la classification par fusion SVM par rapport aux précisions individuelles des classifications SVM mono-sources.

Donnée	OA	Kappa
Optique + NIR	67%	0.62
radar	74%	0.71
MNE	66%	0.61
fusion SVM	94%	0.93

tableau 32 Précisions des classifications mono-sources et fusion SVM - Moorea

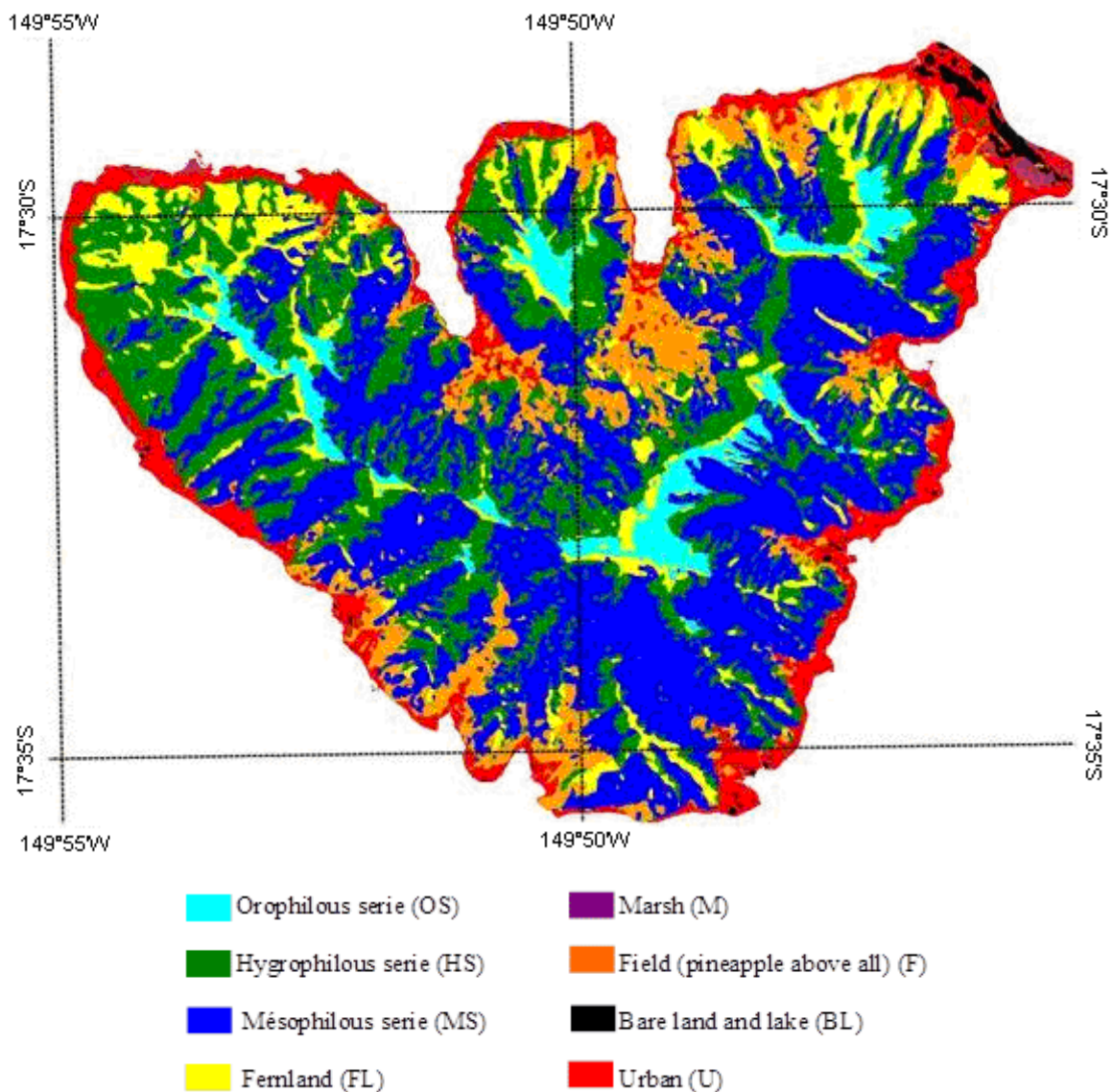


figure 46 Classification par fusion SVM multi-sources - Moorea

Références : (Pouteau et al. psi2008; Pouteau et al. igarss2010a; Pouteau et al. ipt2010)

2. Fusion SVM sélective

L'un des principaux intérêts de la technique de fusion SVM multi-sources réside dans sa capacité à reconnaître un schéma d'erreurs dans un jeu de classifications mono-sources et par apprentissage de classifier correctement une classe n'ayant pas *a priori* des caractéristiques suffisamment discriminantes sur les sources individuelles.

Toutefois, cette technique de fusion ne bénéficie pas à toutes les classes. En effet, certaines possèdent une signature particulière sur un des types de données et sont donc correctement classifiées par une classification mono-source. Dans ce cas de figure, la fusion SVM dégrade la précision de cette classe.

Ainsi, comme certaines classes bénéficient de la fusion SVM alors que d'autres voient leur précision dégradée, une amélioration de l'algorithme de fusion SVM pourrait être de n'appliquer celle-ci qu'aux classes qui en tireront un bénéfice.

Cette technique a été appliquée à grande échelle sur l'île de Moorea, dont j'ai été chargé de produire une carte du milieu naturel dans le cadre du projet Moorea Biocode¹⁹. Ce projet a pour but d'effectuer un inventaire exhaustif de toute forme de vie non-microbienne marine et terrestre de l'île de Moorea.

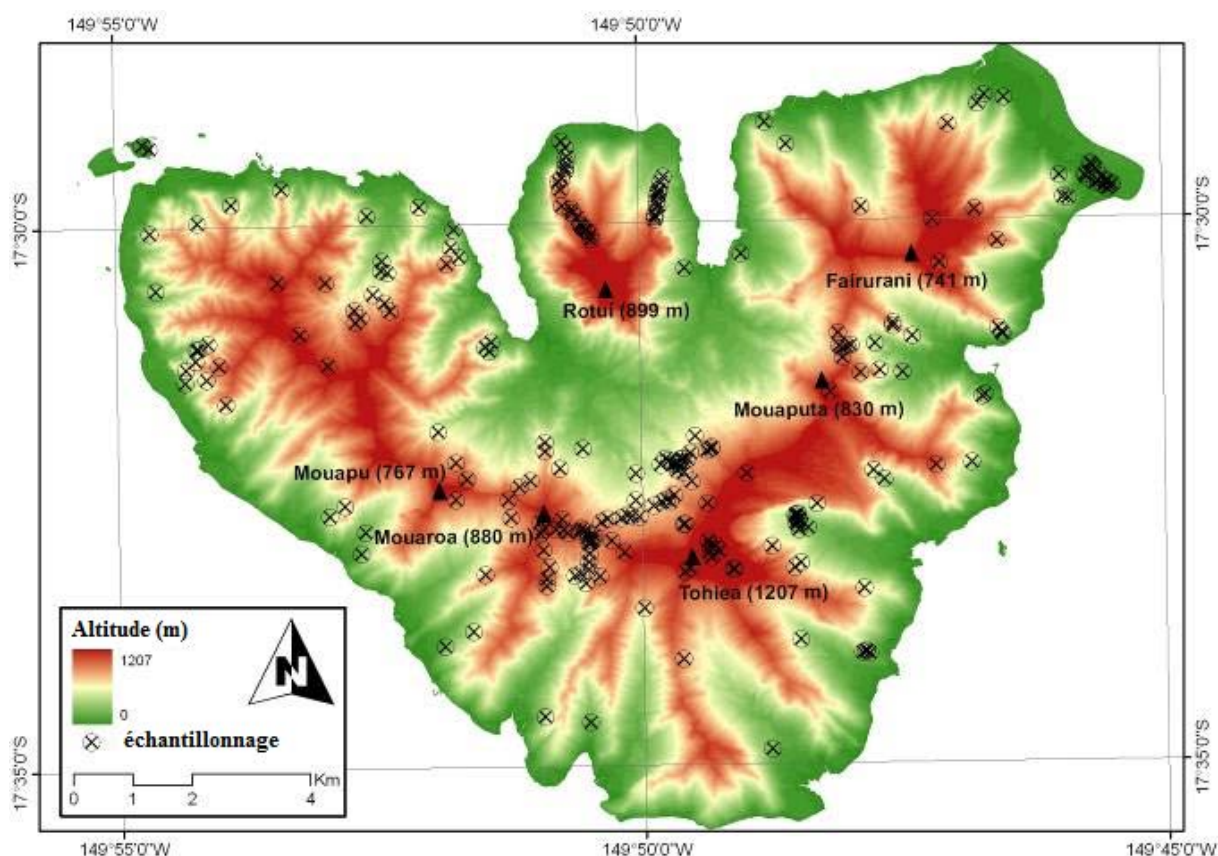


figure 47 Localisation des échantillonnages terrain - Moorea

Le jeu de données disponible est composé d'une donnée radar polarimétrique JPL-AirSAR en bande C et L avec une précision de 5 mètres, d'une donnée optique Quickbird (R, V, B, IR) à 60 centimètres et d'une donnée radar TerraSAR-X en bande X à 2,75 mètres.

¹⁹ <http://mooreabiocode.org/>

Fusion Sélective

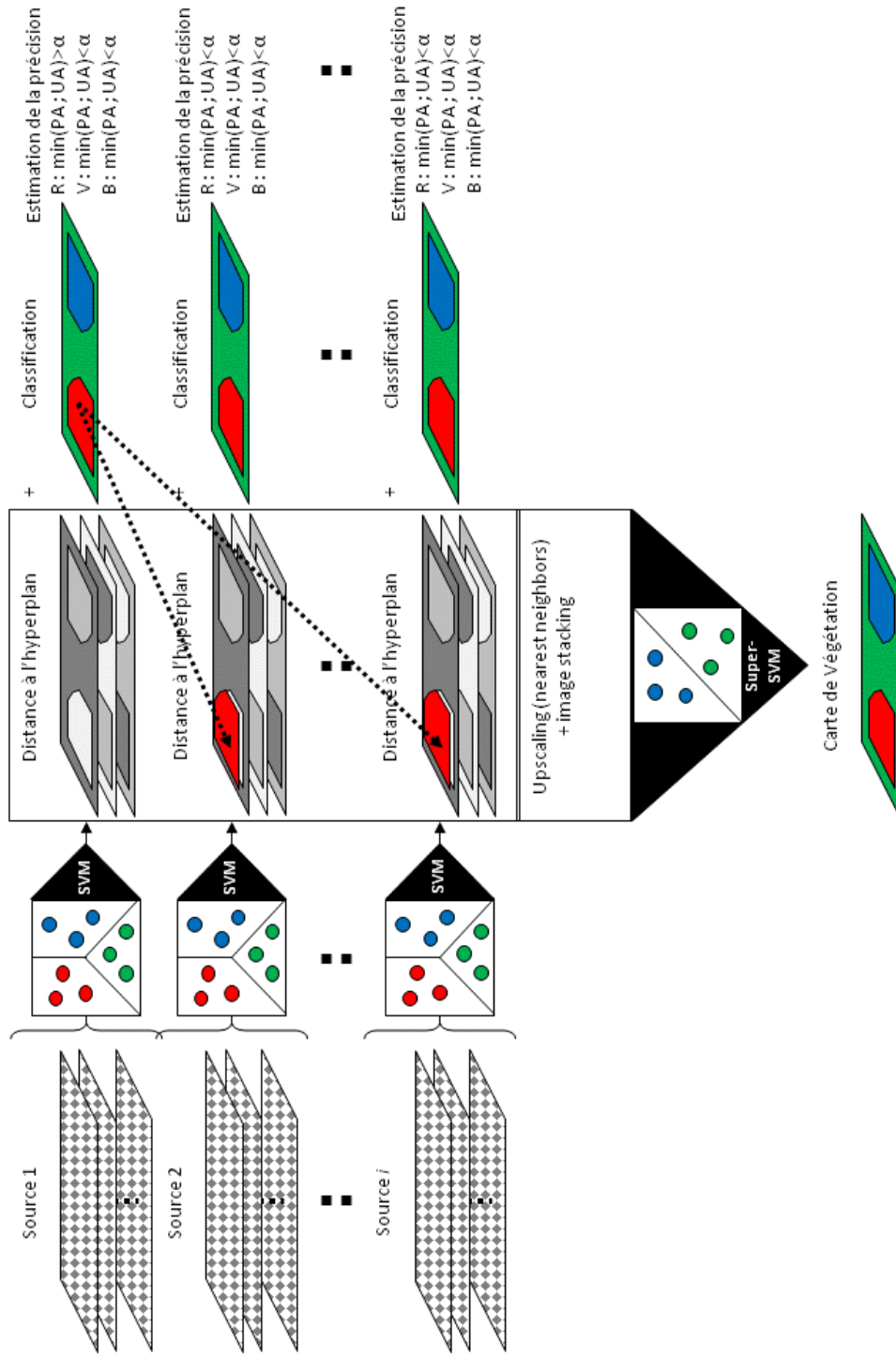


figure 48 Principe de la fusion sélective

Une campagne de vérité terrain a été conduite sur toute l'île de Moorea (à l'exception des zones anthropisées), la *figure 47* illustre la localisation des échantillonnages terrain.

La technique proposée consiste à effectuer dans un premiers temps quatre classifications mono-sources sur : la donnée optique, les attributs de textures optiques, la donnée radar et les données ancillaires tirées du *MNE*. Les précisions globales sont reportées dans le *tableau 33*.

Sources	OA	Kappa
1. Donnée optique multi-spectrale	62.7%	0,6
2. Donnée optique texturale	50.3%	0,47
3. Donnée SAR	36.7%	0,33
4. Donnée auxiliaire tirée du MNT	64.2%	0,62

tableau 33 Précisions des classifications mono sources - Moorea

Les matrices de confusions des classifications mono-sources ont été analysées pour déterminer les classes qui bénéficient d'une bonne séparabilité dans le cas mono-source.

Un paramètre de seuil α permet de définir si une classe est bien classifiée (sur une ou plusieurs sources), ou si elle est en difficulté (*i.e.* si sa précision n'est suffisante sur aucune source classifiée individuellement).

- Lorsqu'une classe est considérée en difficulté, la fusion *SVM* est appliquée à cette classe afin d'en améliorer la précision. La *figure 48* illustre le schéma de fusion sélective proposé.
- Dans le cas où une classe est suffisamment bien classifiée (si sa précision est supérieure au seuil α) et n'obtiendrait aucun bénéfice de la fusion, nous n'opérerons pas la fusion et garderons les pixels concernés pour la classification finale en les excluant de la phase de fusion.
- Dans le cas où une classe est au-dessus du seuil dans plusieurs résultats de classifications mono-sources, nous choisirons les résultats de la classification qui obtiendra la meilleure précision.

La seconde étape de fusion *SVM* est opérée uniquement pour les classes ayant une précision inférieure au seuil α .

Cette technique a été dénommée « fusion sélective (*SELF*) » ou « fusion pour les classes en difficulté ».

Le *tableau 34* reporte les précisions individuelles par classe de la classification finale obtenue par fusion sélective. La colonne **source** indique l'origine (*cf. tableau 33* pour 1, 2, 3 et 4 et F pour fusion *SVM*).

Certaines classes ne bénéficient que d'une seule source :

- Les classes *Pinus caribaea*, *Falcataria moluccana*, *Casuarina equisetifolia*, *Dicranopteris linearis* et *Miscanthus floridulus* sont bien classifiées uniquement à partir de la donnée optique.
- La classe *Aleurites moluccana* est obtenue à partir des attributs de textures optiques.
- La classe *Hibiscus tiliaceus* est obtenue à partir de la donnée radar.
- Les classes *Typha domingensis*, *Montane cloud forest* et *Summit shrubland* sont obtenus à partir des indices géomorphologiques tirés du *MNE*.

Toutes les autres classes bénéficient de la fusion *SVM* et améliorent ainsi leurs précisions de classification.

Classes	Source	PA(%)	UA(%)
Plantations			
Pinus caribaea	1	72.1	67.1
Falcataria moluccana (coral)	1	79.0	46.1
Cocos nucifera	F	90.4	90.4
Coastal vegetation			
Typha domingensis	4	100	100
Low- to mid-elevation mesic to moist vegetation			
Metrosideros collina (dark green)	F	83.1	86.3
Casuarina equisetifolia	1	90.1	91.0
Dicranopteris linearis	1	98.2	93.0
Leucaena leucocephala	F	80.3	79.2
Syzygium cumini	F	73.9	60.7
Miscanthus floridulus	1	86.0	88.4
Low- to mid-elevation moist to wet vegetation			
Neonauclea forsteri (magenta)	F	61.7	41.3
Aleurites moluccana (white)	2	92.5	86.5
Inocarpus fagifer	F	50.0	39.4
High-elevation vegetation			
Montane cloud forest (cyan)	4	95.3	96.4
Summit shrubland	4	98.9	94.8
Ubiquitous			
Hibiscus tiliaceus (green)	3	24.6	47.3
Spathodea campanulata	F	55.2	70.0

tableau 34 Précisions par classe de la fusion sélective SVM

Une nette amélioration de la précision globale de la fusion SVM sélective (*SELF*) est obtenue par rapport à la fusion SVM systématique (tableau 35).

Méthode	OA	Kappa
Fusion SVM systématique	67.0%	0,65
Fusion SVM sélective (<i>SELF</i>)	78.2%	0,77

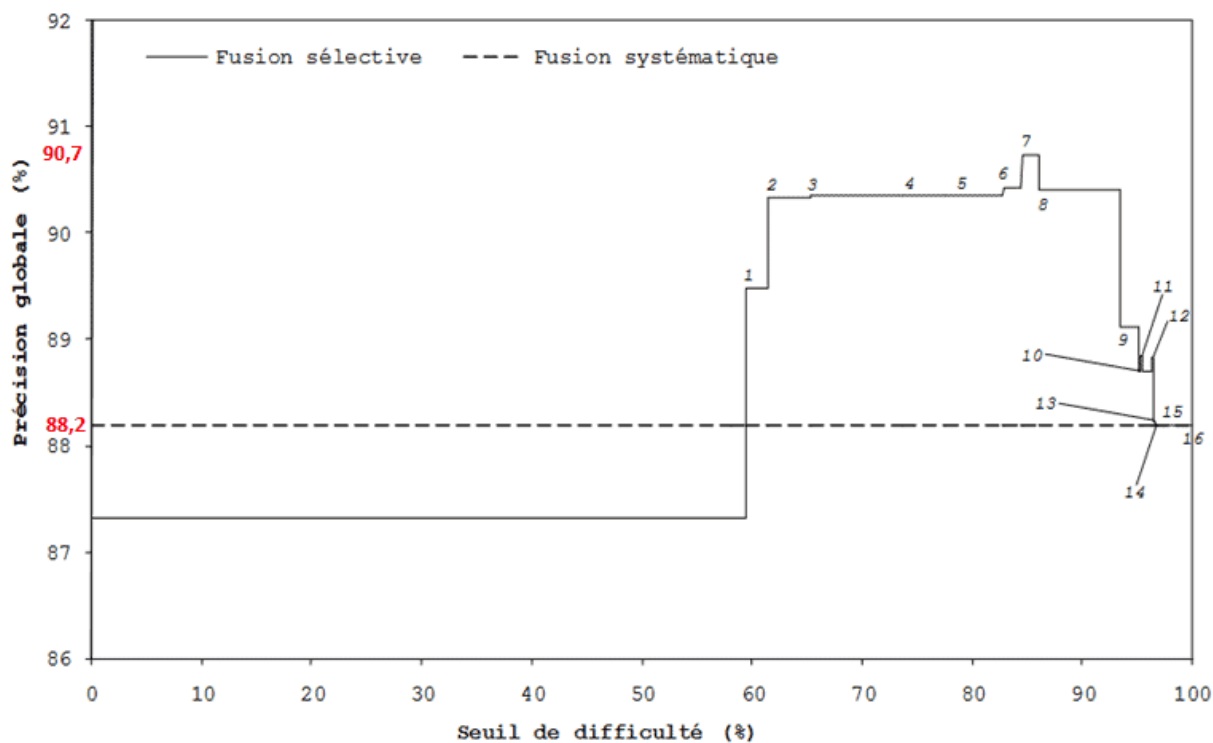
tableau 35 Précisions de la fusion SVM vs fusion SVM sélective

La méthode proposée permet de réduire considérablement les temps de calcul par rapport à une fusion SVM classique.

La *figure 49* montre les résultats de la précision globale de la classification finale lorsque l'on fait varier le seuil α .

Lorsque le seuil est en dessous de 60, toutes les classes sont exclues de la fusion, la précision est de 87,3%. Lorsque l'on augmente le seuil, les classes sont incluses progressivement dans la fusion, augmentant petit à petit la précision globale de la classification. A l'extrême, toutes les classes sont fusionnées avec un seuil de 100, la précision devient égale à celle de la fusion systématique.

Notons que lorsque le seuil est compris entre 60 et 96, la précision dépasse celle de la fusion SVM simple, c'est là que la fusion SVM sélective trouve son intérêt. La meilleure précision est obtenue pour un seuil compris entre 86 et 87 et une précision de 90,7%.



- | | |
|---|------------------------------------|
| 1 : Forêts anthropiques + forêts à <i>Neonauclea</i> | 9 : Faciès à <i>Aleurites</i> |
| 2 : Ripisylves à <i>Inocarpus</i> | 10 : Brousses à <i>Leucaena</i> |
| 3 : Matrice à <i>Hibiscus</i> | 11 : Forêts ombrophiles |
| 4 : Faciès à <i>Spathodea</i> | 12 : Landes à <i>Dicranopteris</i> |
| 5 : Forêts à <i>Metrosideros</i> | 13 : Végétation subalpine |
| 6 : Plantations de <i>Cocos</i> | 14 : Plantations de <i>Pinus</i> |
| 7 : Bosquets à <i>Falcataria</i> | 15 : Bosquets à <i>Casuarina</i> |
| 8 : Savanes à <i>Miscanthus</i> , <i>Melinis</i> , <i>Psidium</i> et <i>Dodonea</i> | 16 : Marécages à <i>Typha</i> |

figure 49 Précision globale en fonction du seuil α

La figure 50 illustre la carte des milieux naturels de l'île de Moorea obtenue par fusion sélective (*SELF*) de données Quickbird, AirSAR, TerraSAR-X et ancillaire.

Références : (Pouteau, Lardeux et al. *isrse2011*; Pouteau et al. *igarss2011*; Pouteau et al. *jstars2012*).

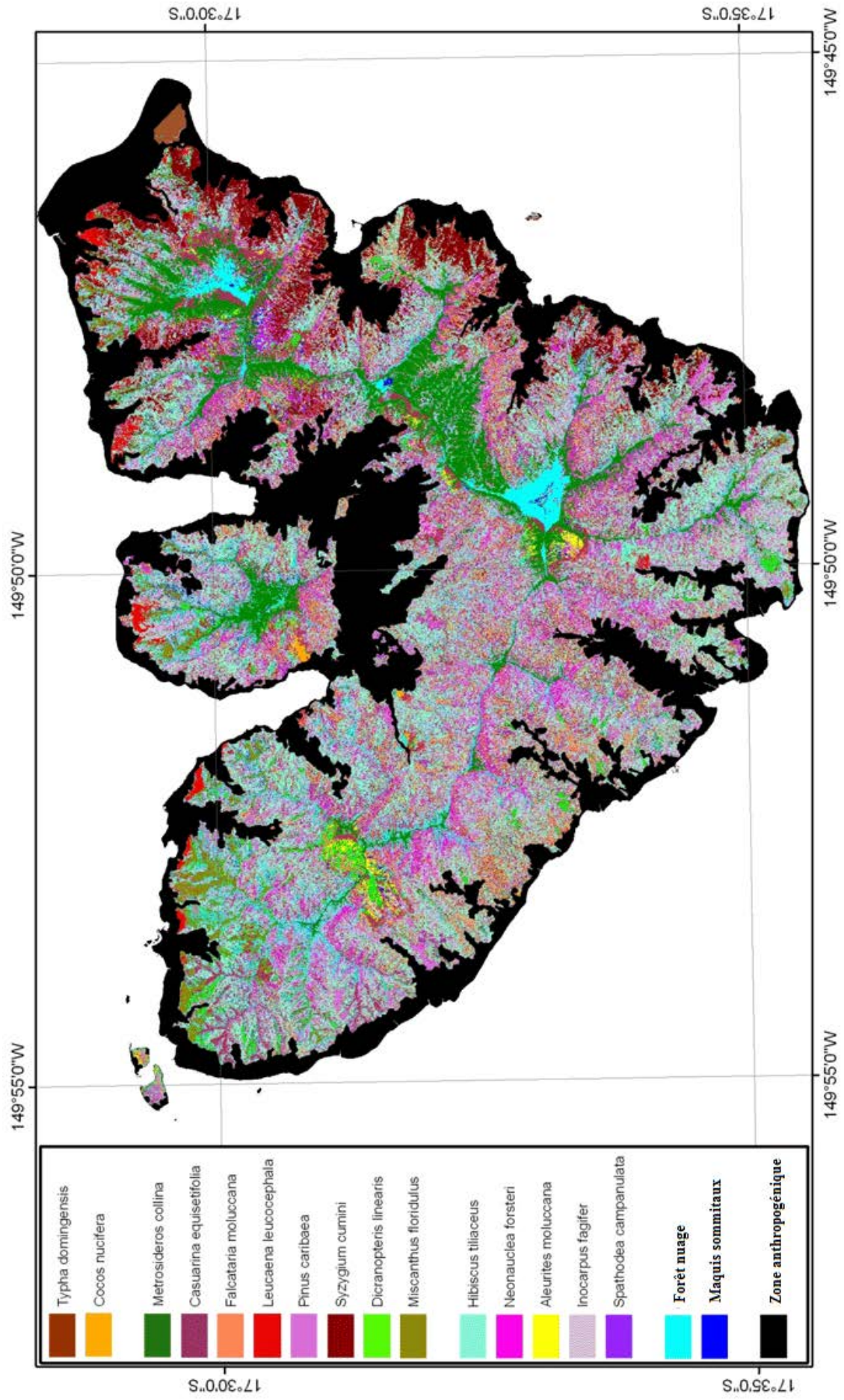


figure 50 Carte des milieux naturels de Moorea obtenue par fusion SVM sélective

B. ESPECES ENVAHISSANTES ET ESPECES RARES

1. Cartographie de *Miconia calvescens*

Le chapitre II.C.2 a montré la capacité de l'algorithme SVM à modéliser la niche écologique de *Miconia* par la classification supervisée de cinq descripteurs biophysiques abiotiques. Une autre étude sur la cartographie de *Miconia* a été conduite sur l'île de Tahiti (vallée de la Papenoo).

Seules les variables géomorphologiques suivantes ont été considérées : l'élévation, les précipitations, la pente, l'insolation annuelle potentielle, l'exposition, le *CTI*. À l'échelle de la vallée de la Papenoo, l'influence de l'exposition aux vents dominants a été considérée comme négligeable.

Une carte du couvert végétal a été ajoutée à ces descripteurs biophysiques. Cette carte a été obtenue par classification de la donnée Quickbird additionnée de ses attributs de texture. Ce facteur biotique est important car la couche supérieure de la végétation a aussi une influence sur l'invasion du *Miconia* ; par exemple : nous ne trouvons pas de *Miconia* sous les forêts de Bambous.

La figure 51 propose un schéma synthétique de la procédure proposée.

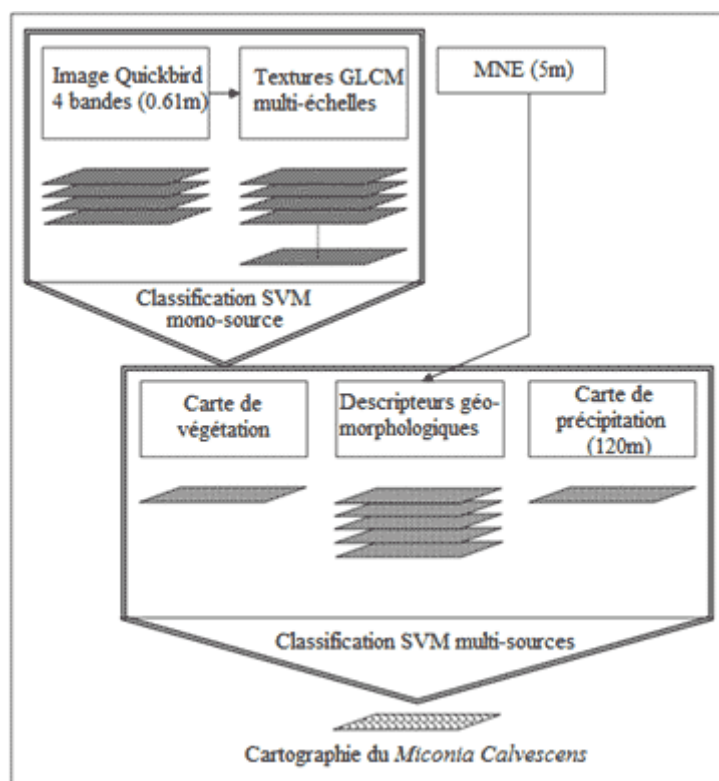


figure 51 Chaîne de traitement pour la cartographie du *Miconia*

Cette méthode a permis d'obtenir une cartographie du *Miconia* sur la vallée de la Papenoo (figure 52). Dans cette vallée, comme dans toute l'île de Tahiti, les biologistes considèrent que le *Miconia* est arrivé à son expansion maximum. Cette carte de l'habitat potentiel est donc équivalente à l'invasion définitive du *Miconia*. Cette cartographie est la première cartographie du *Miconia* par des méthodes de télédétection.

Références : (Pouteau et al. ecomod2011).

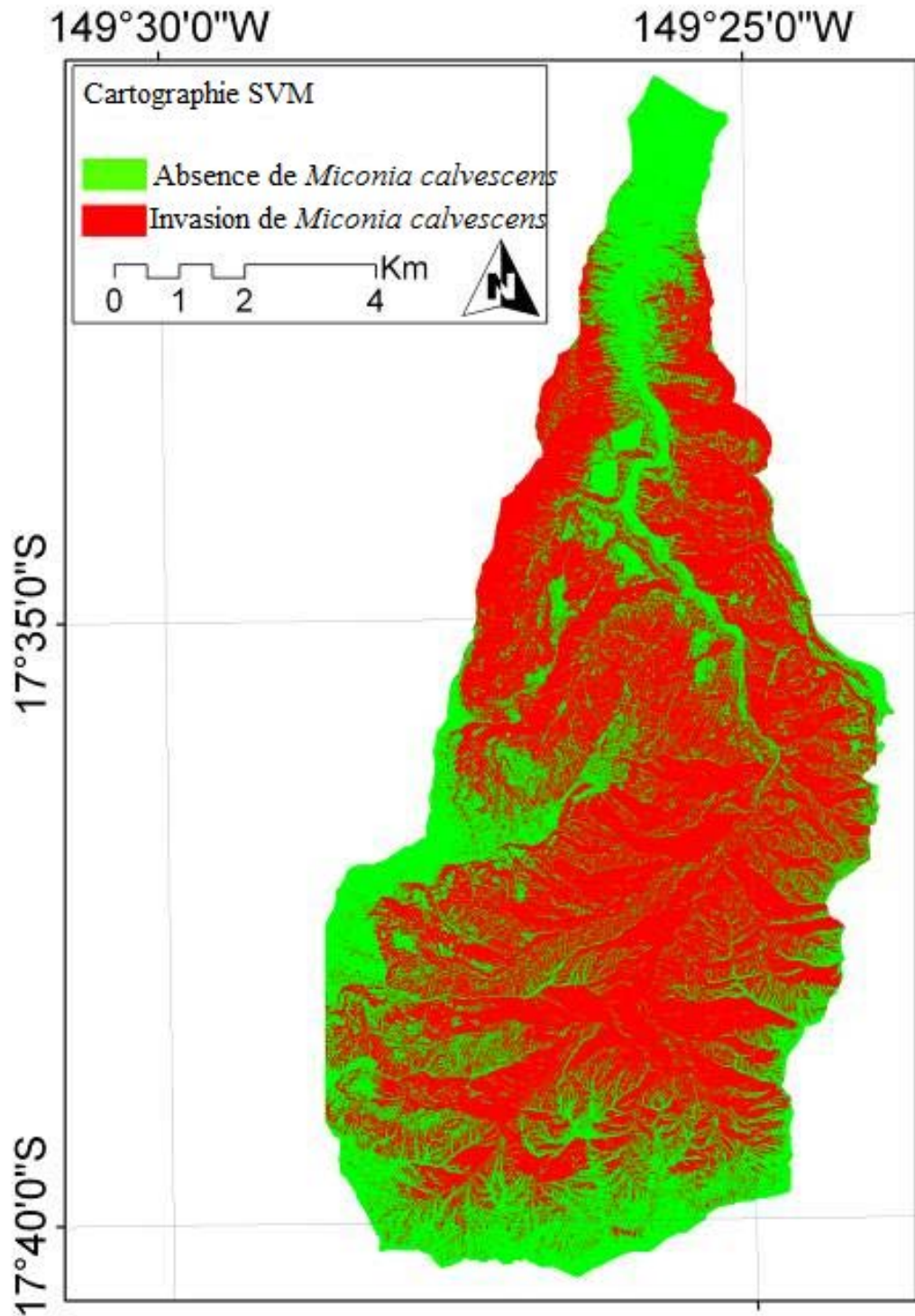


figure 52 Cartographie SVM du Miconia - vallée de la Papenoo - Tahiti

2. Niches écologiques d'espèces rares

Les îles océaniques ont la particularité d'avoir un biotope unique avec un fort taux d'endémisme mais aussi une forte sensibilité aux perturbations anthropogéniques causant une diminution du nombre et de la distribution de certaines espèces.

Ces espèces rares sont souvent regroupées sur une petite quantité de stations dans quelques milieux préservés. Il en résulte un faible nombre d'échantillons disponibles, ce qui pose un problème pour la plupart des algorithmes de modélisation de niches écologiques.

L'une des propriétés intéressantes de l'algorithme *SVM* est son besoin réduit en pixels d'entraînement pour définir correctement la frontière entre deux classes. Son utilisation semble donc particulièrement adaptée à la modélisation de niches écologiques d'espèces rares.

Pour valider cette hypothèse, les résultats obtenus par l'algorithme *SVM* sont comparés avec ceux obtenus par un autre algorithme de type réseaux de neurones, l'algorithme *Random Forest (RF)*. Ce dernier a obtenu lors de plusieurs études comparatives récentes les meilleurs résultats pour prédire la distribution d'espèces rares.

Le schéma adopté est celui décrit dans la *figure 53* suivante.

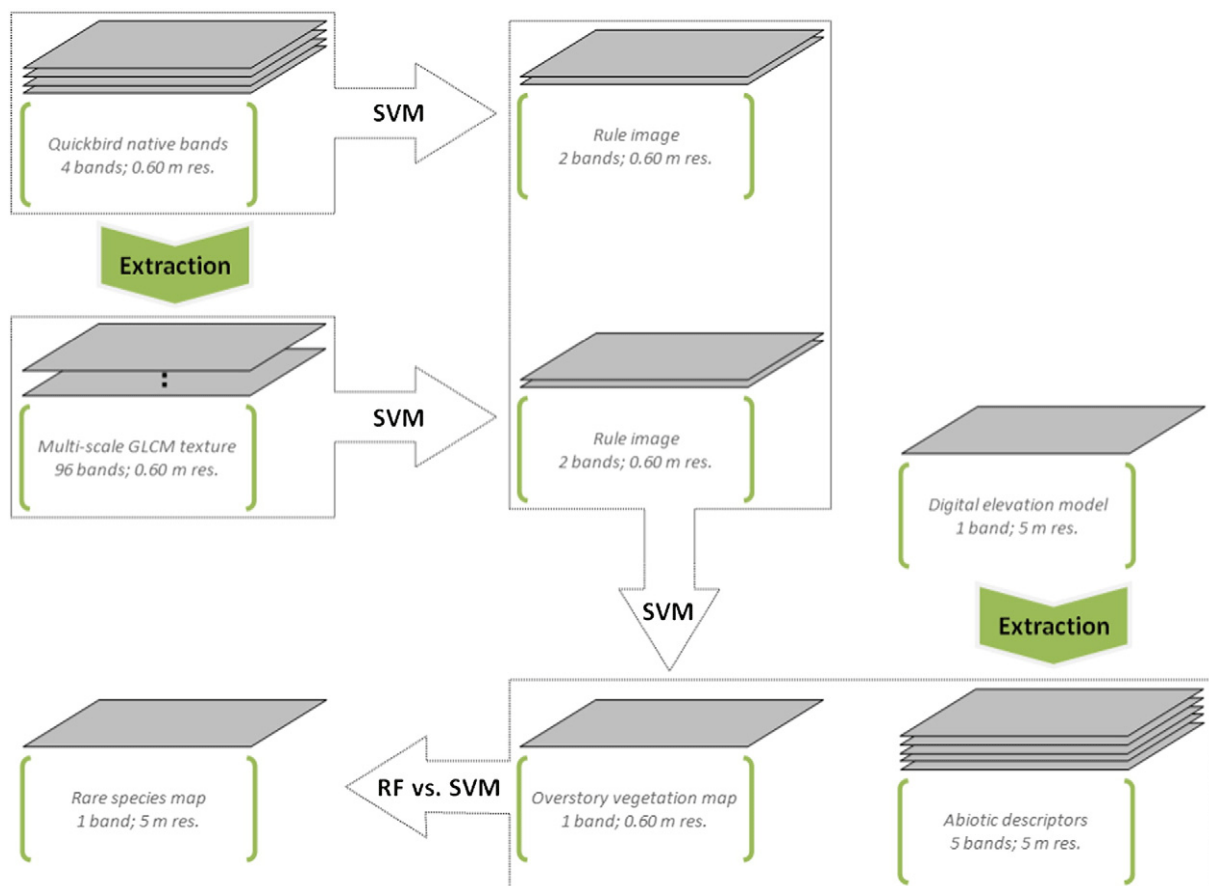


figure 53 Schéma de la comparaison des algorithmes Random Forest et SVM

Cinq descripteurs abiotiques ont été produits à partir d'un modèle numérique d'élévation. Une carte de la végétation dominante (*figure 54*), a été produite en fusionnant les résultats de la classification *SVM* des 4 bandes d'une image Quickbird et de ses attributs de textures.

La comparaison a été effectuée sur la capacité des algorithmes *RF* et *SVM* à prédire l'habitat potentiel des espèces rares à partir de ces deux types d'informations complémentaires.

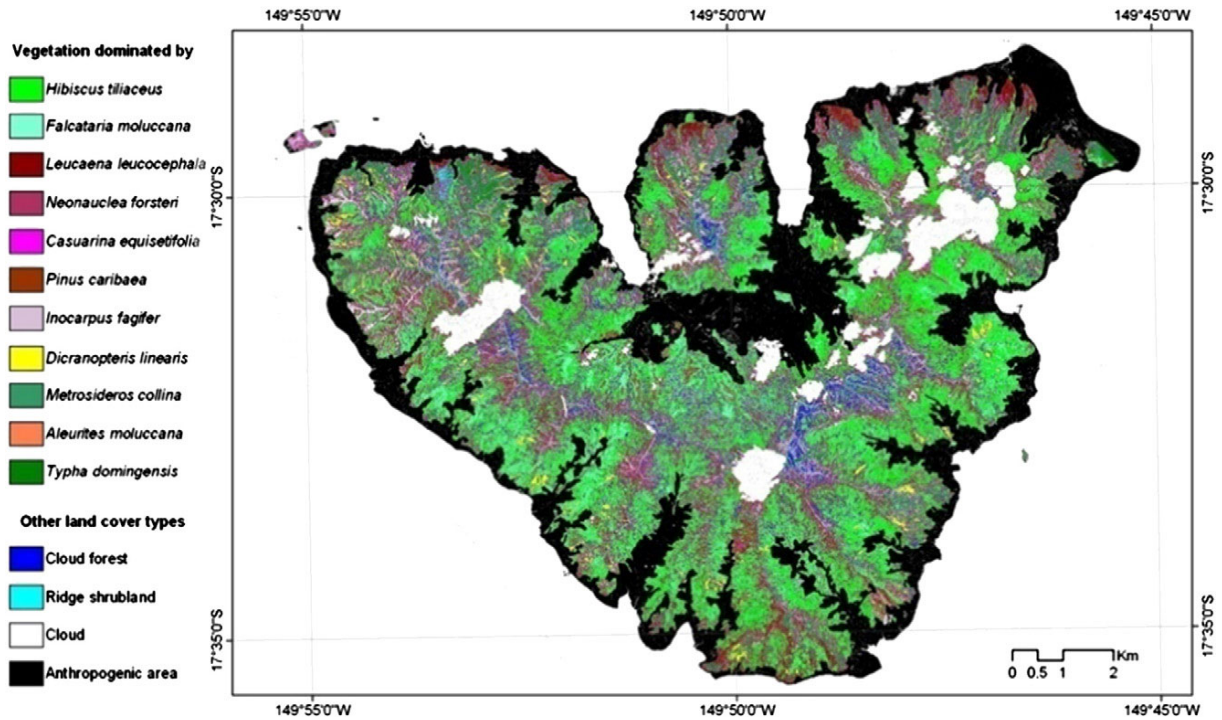


figure 54 Carte de végétation de Moorea par classification SVM d'image Quickbird

Cette étude a porté sur 3 espèces rares présentes sur l'île de Moorea, que sont *Lepinia taitensis* (figure 55-a) dont 28 occurrences sont connues, *Pouteria tahitensis* (figure 55-b) dont 20 occurrences sont connues, et *Santalum insulare* (figure 55-c) dont 81 occurrences sont connues.

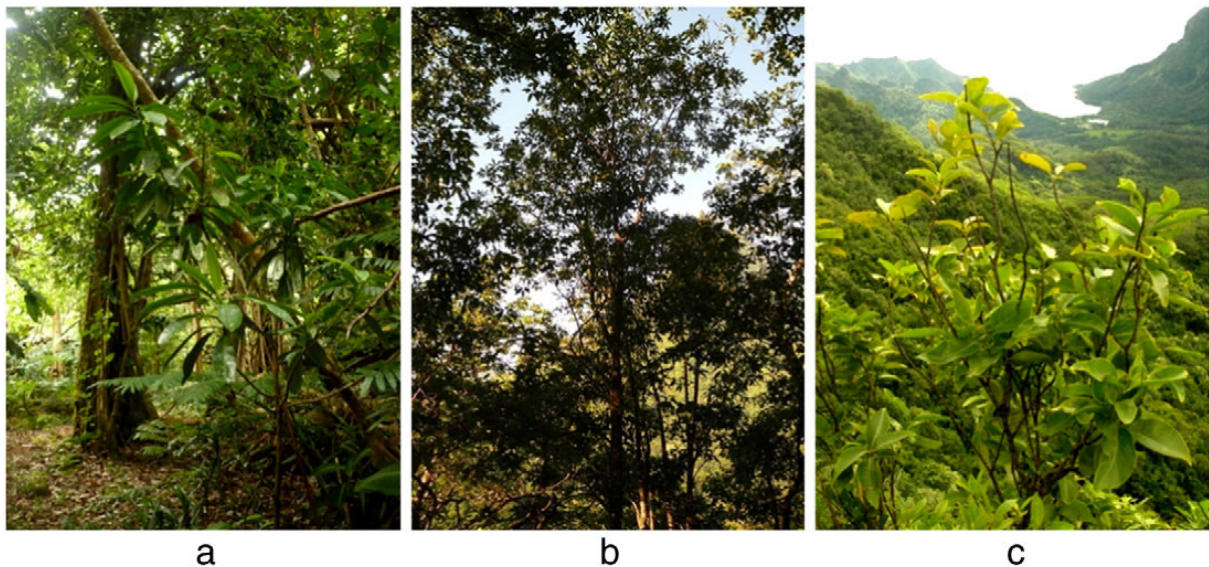


figure 55 Espèces rares : a) *Lepinia taitensis*, b) *Pouteria tahitensis*, c) *Santalum insulare*

La figure 56 présente les stations d'échantillonnages terrain pour la cartographie de la végétation et pour la localisation des occurrences des trois espèces rares considérées.

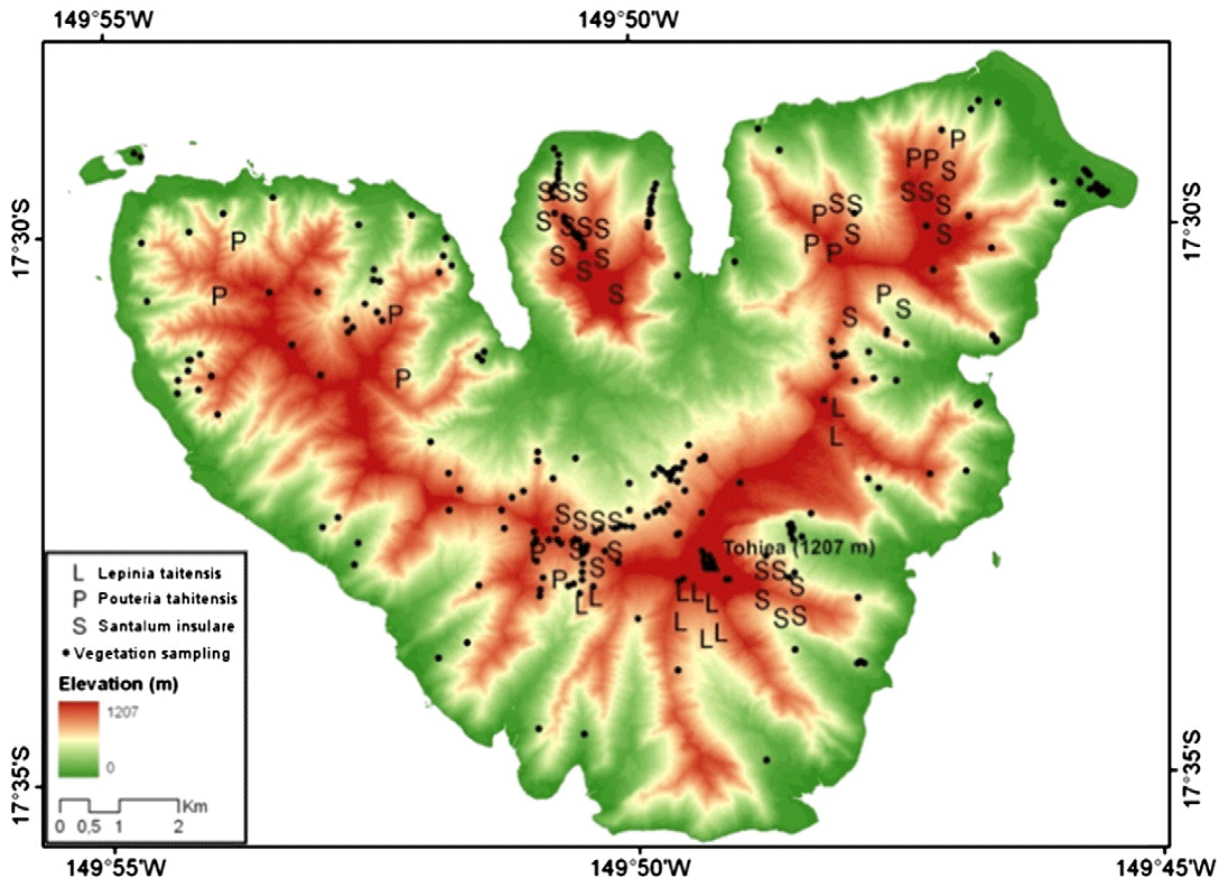


figure 56 Échantillonnages terrains

L'apprentissage a été effectué sur 67% des échantillons de la classe « présence » pour les trois espèces, l'évaluation des résultats a été effectuée sur les 33% restants. Les pixels représentant la classe « absence » (qui sont en fait des « pseudo-absences ») sont choisis au hasard dans les zones non-échantillonnées et non susceptibles de contenir les espèces concernées.

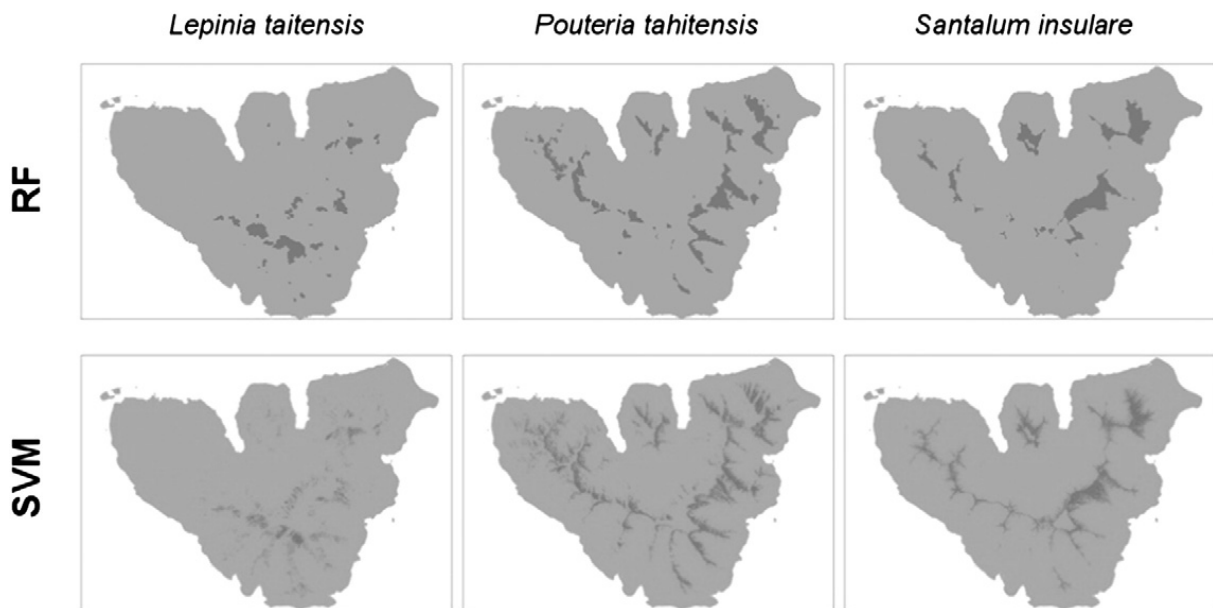


figure 57 Cartographie RF et SVM de l'habitat potentiel de 3 espèces rares - Moorea

La figure 57 compare les résultats des cartographies des niches écologiques des 3 espèces rares effectuées à l'aide de l'algorithme SVM et de l'algorithme RF. Les résultats sont

relativement similaires mais l'algorithme *RF* a tendance à surestimer les surfaces par rapport aux résultats de l'algorithme *SVM*. Ce dernier montre des contours plus fins et semble plus sensible à la microtopographie.

En termes de précision obtenue, les deux algorithmes obtiennent de très bons résultats avec une précision globale comprise entre 87 et 97%.

Si l'on regarde classe par classe, les algorithmes obtiennent des résultats similaires pour les classes *Lepinia Taitensis* et *Santalum insulare*. Cependant pour la classe *Pouteria Tahitensis*, ayant le plus faible nombre d'occurrences, l'algorithme *SVM* obtient de meilleurs résultats. Ce qui validerait la thèse émise en début de paragraphe.

Références : (Pouteau et al. *ecoinfo*2012).

C. SANTE DU RECIF CORALLIEN

Cette dernière application est un exemple d'utilisation conjointe de données issues des SIG et de la télédétection.

L'île de Moorea, située à une demi-heure de bateau de l'île de Tahiti, est de plus en plus prisée par les personnes travaillant à Papeete qui choisissent de vivre sur Moorea, préférant les navettes en bateau aux embouteillages biquotidiens pour accéder à la ville. Moorea connaît de ce fait une poussée démographique rapide qui accroît la pression anthropique sur le récif corallien notamment.

Le corail est soumis à deux types de pressions d'origine anthropique : les menaces distales comme le réchauffement planétaire, l'acidification des océans, etc. et les menaces proximales comme la surpêche lagonnaire, la pollution terrestre, l'apport terrigène dans le lagon etc.

Le récif barrière est relativement isolé de l'influence de facteurs terrestres par le chenal et ses courants puissants, il sera donc essentiellement affecté par des facteurs distaux peu aisés à modéliser. Le récif frangeant en revanche est directement influencé par l'ensemble des facteurs proximaux provenant du bassin versant qui le surplombe, c'est donc sur ce récif frangeant que se focalise cette étude.

L'objet de celle-ci est de modéliser l'état de santé du corail en fonction de facteurs socio-écologiques liés au bassin versant. Il sera ainsi possible de vérifier quels facteurs ont le plus d'influence sur la santé du récif frangeant.

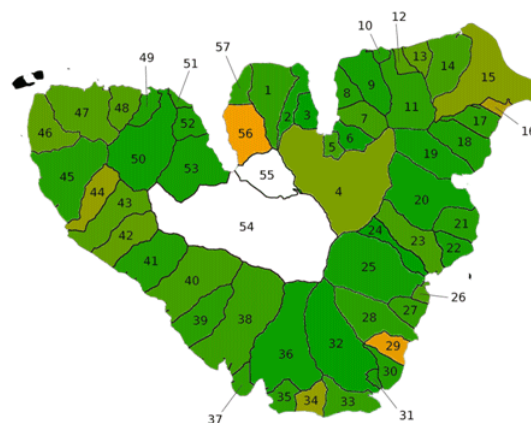


figure 58 Bassins versants de Moorea

Un modèle numérique d'élévation à 5 mètres permet de découper l'île de Moorea en 57 bassins versants dont seuls 55 seront conservés. En effet, la baie d'Opunohu ne possède pas de récif frangeant et les deux bassins versants concernés sont exclus de la modélisation. Une

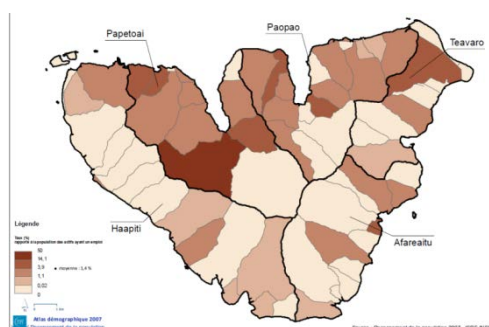
zone d'influence est délimitée pour chaque bassin versant. Elle correspond au prolongement de ce dernier sur le récif frangeant et subira les apports du bassin versant correspondant.

L'école pratique des hautes études (EPHE-CRIOBE) a effectué une classification SVM du corail de l'île de Moorea²⁰ à partir d'une image Quickbird de 2006. Cette classification sert de base au calcul d'un indice de santé du corail (*Reef Health Index - RHI - Eq.4*) pour chacune des 55 zones d'influences.

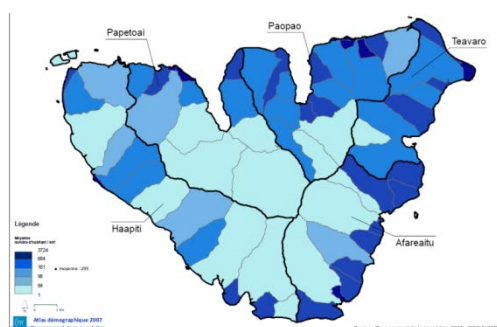
$$RHI = \frac{\text{Vivant} - \text{Mort}}{\text{Vivant} + \text{Mort}} \quad (\text{Eq } 4)$$

Cet indice tend vers 1 lorsque la couverture en corail vivant prédomine et tend vers 0 lorsque la couverture du corail mort prédomine, il est un indicateur de l'état de santé du corail.

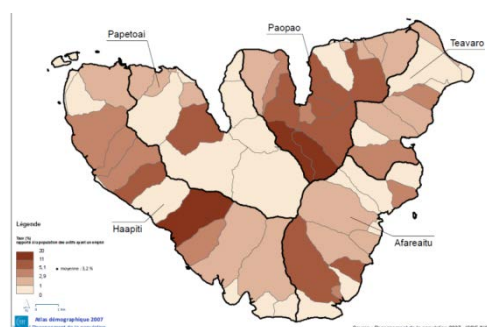
Pour modéliser cet indice *RHI*, nous avons accès à trois types de facteurs socio-écologique : Les facteurs sociologiques, les facteurs physiographiques, et les facteurs d'occupation des sols.



a) Taux de fermiers



b) Densité de population



c) Taux de pêcheurs

figure 59 Variables sociologiques de Moorea

Les facteurs sociologiques proviennent de l'institut de la statistique de la Polynésie française²¹ sous la forme de couches SIG. Ces données sont le taux de fermiers par rapport à la population active, le taux de pêcheurs par rapport à la population active et la densité de population. Elles sont découpées par bassin versant (figure 59).

²⁰ Collin, A., P. Archambault, and S. Planes, "Revealing coral reefs' states with a versatile continuous spatial model", *Ecological applications*, submitted, 2013.

²¹ ISPF (Institut de la Statistique de Polynésie française), "Atlas démographique de la Polynésie française : recensement de la population 2007", *Regards special issue 2*, 2007.

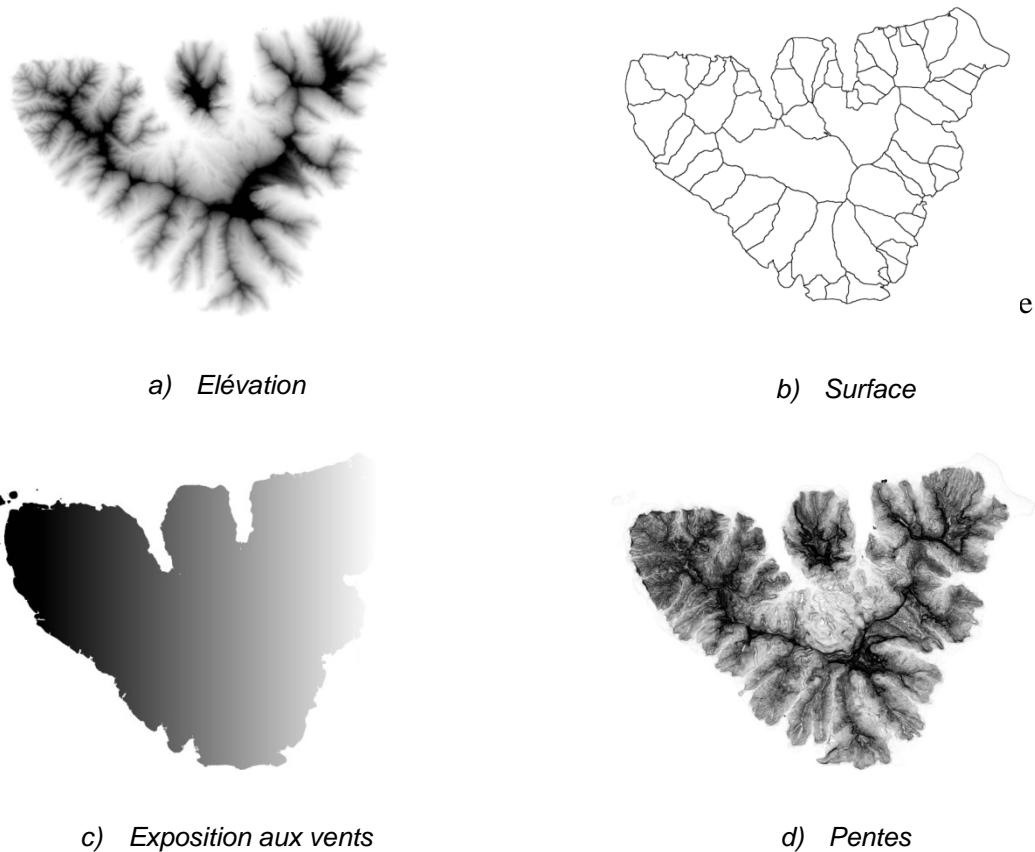


figure 60 Variables physiographiques de Moorea

Les facteurs physiographiques sont la moyenne et l'écart type de l'élévation et de la pente, l'exposition aux vents ainsi que la surface totale de chaque bassin versant (figure 60).

Enfin les facteurs d'occupation des sols sont issus de la carte des milieux naturels de Moorea (figure 50) regroupés selon 8 classes d'occupation des sols : zones construites, zones cultivées, zones herbacées, zones arbustives, zones forestières, zones à plante invasives, zones à plantes natives et zones à plantes non natives.

La modélisation est effectuée par des régressions successives opérées avec l'algorithme de *Boosted Regression Tree*. Il est calibré à partir de 37 bassins versants tirés au hasard, et validés sur les 18 restants.

Une fois le modèle validé, l'influence relative de chaque facteur est calculée sur le modèle complet, ce qui permet de mettre en valeur les facteurs qui ont une forte influence sur la santé du corail.

Ce modèle atteint une précision de 0.74 (*Area Under the ROC Curve*²²), ce qui correspond à une modélisation « correcte ».

Les variations que ce modèle n'explique pas pourraient être causées par une précision insuffisante des données à l'entrée du modèle (comme la cartographie du corail ou de l'occupation des sols), mais aussi par l'influence de variables distales comme le réchauffement, et l'acidification des océans qui ne sont pas prises en compte dans ce modèle.

²² Andrew P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recognition*, Volume 30, Issue 7, July 1997, Pages 1145-1159, ISSN 0031-3203

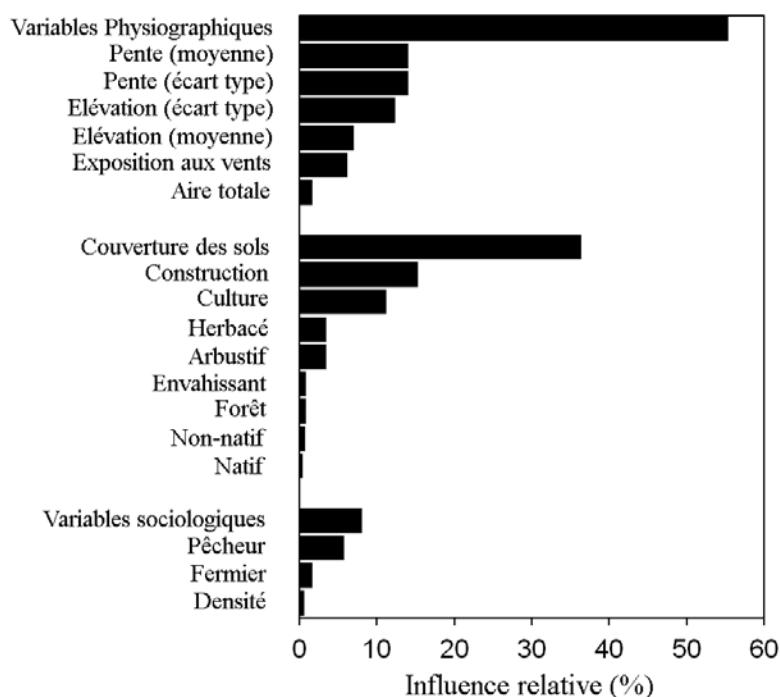


figure 61 Influence relative de chaque variable sur la santé du corail (RHI)

Nous pouvons néanmoins analyser ces résultats en détail : les groupes de variables ayant la plus grande influence sur l'indice *RHI* et donc sur la santé du corail frangeant sont les variables physiographiques en premier lieu, viennent ensuite les variables liées à la couverture des sols puis les variables sociologiques (figure 61).

Les facteurs de pente et d'élévation ont une grande influence. Ces variables sont liées à la vulnérabilité des sols à l'érosion et donc à l'apport terrigène au lagon. Le corail est en bonne santé lorsque la pente est comprise entre 20° et 30°. Cependant, il est difficile de décorréliser ce facteur de l'activité anthropique ; en effet, les activités humaines sont situées dans des zones de faible pente, ce qui pourrait expliquer pourquoi le *RHI* est faible pour les pentes inférieures à 20° (au-delà de 30°, on peut supposer que la pente induit un fort apport terrigène lors des fortes pluies).

L'influence de l'activité humaine sur la santé du corail frangeant est confirmée par la surface de zones construites (variable ayant le plus d'influence sur l'indice *RHI*). Lorsque celle-ci est inférieure à 20ha, le récif frangeant est en bonne santé. Au-delà de 20ha, la valeur du *RHI* baisse drastiquement.

Cette variable n'est pas linéairement corrélée avec la densité de population, qui a de son côté une très faible influence sur le *RHI*, ce qui tendrait à démontrer que ce sont les constructions comme les hôtels, zones industrielles, ports, aéroports, toutes situées en zones non-résidentielles qui auraient la plus grande influence sur la santé du corail.

Cette étude est un bon exemple de la modélisation de l'influence d'un certain nombre de facteurs sociétaux et environnementaux sur la santé d'un milieu naturel à partir de données SIG et de télédétection. Ce type de modélisation permet de mieux comprendre la relation entre le milieu naturel et des facteurs facilement mesurables et gérables. Il constitue ainsi un outil précieux pour les décideurs publics, leur permettant de mettre en place une politique raisonnée préservant au mieux l'environnement.

Références : (Pouteau et al. 2013)

V. VISION SYNOPTIQUE DE MON DOMAINE DE RECHERCHE

A. SYNTHÈSE

1. Principaux résultats

Mon objectif principal est la cartographie du milieu naturel terrestre des îles hautes. Or, l'élaboration de cette cartographie s'avère délicate du fait de la complexité et l'hétérogénéité de la distribution spatiale de la végétation, causées par les contraintes fortes liées à la topographie abrupte, la taille réduite des îles, mais aussi l'activité anthropique.

La cartographie la plus aboutie a été obtenue en intégrant conjointement un maximum d'informations complémentaires: informations spectrales et texturales à partir de la donnée optique, informations structurelles et polarimétriques à partir de la donnée radar, informations géomorphologiques issues du *MNE* et des cartes météorologiques. Ces dernières apportent une information complémentaire sur la distribution spatiale des espèces.

Dans ce contexte de cartographie très ardu, l'utilisation de l'algorithme *SVM* s'est révélée particulièrement appropriée du fait de ses propriétés intrinsèques que sont : sa nature non-descriptive, sa capacité de discrimination basée sur un faible nombre de pixels d'apprentissage, sa capacité à traiter un grand nombre de données hétérogènes,

Ce faisant, des problématiques environnementales importantes ont été abordées telles la cartographie de l'espèce envahissante *Miconia calvescens*, la cartographie de l'habitat d'espèces rares par modélisation de leur niche écologique, la cartographie de la forêt de nuages de Tahiti et sa régression probable liée au réchauffement planétaire, l'étude de la cocoteraie des Tuamotu (ressource naturelle importante pour la Polynésie française), la modélisation de l'état de santé du récif corallien frangeant, etc. Mais la contribution la plus significative a été sans aucun doute la cartographie complète du milieu naturel de Moorea qui a été proposée dans le cadre du projet international [Moorea Biocode](#).

Des résultats plus fondamentaux ont aussi été obtenus comme la validation de l'algorithme *SVM* pour classifier de la donnée radar, mais aussi pour des cas complexes (grand nombre de bandes, données hétérogènes, données à très hautes résolutions), l'étude comparative d'indices polarimétriques radar pour discriminer la végétation polynésienne, une méthode de sous-échantillonnage pour pallier le faible nombre de zones d'apprentissage homogènes, la validation de la méthode d'échantillonnage terrain sur les écotones pour les algorithmes de classification et en particulier pour l'algorithme *SVM*, l'utilisation de l'algorithme de ligne de partage des eaux pour segmenter les couronnes de cocotiers sur les images optiques très haute résolution, l'intérêt de la méthode de fusion *SVM* sur la végétation tropicale, une optimisation d'un schéma de fusion *SVM* connu, etc.

2. Transversalité

Nous l'avons vu, la télédétection est une chaîne de traitements complexe. J'ai abordé chaque maillon de celle-ci afin d'en maîtriser tous les aspects essentiels et acquérir une expertise globale sur la télédétection environnementale. L'intérêt de cette démarche c'est qu'elle ouvre de nombreuses potentialités d'études scientifiques.

De la même manière, j'ai essayé d'être aussi exhaustif que possible sur les choix de mes études tant au niveau de la localisation, des milieux que de la donnée utilisée.

J'ai ainsi travaillé sur les différents types d'îles que l'on peut trouver en Polynésie française:

- îles hautes,

- atolls coralliens.

J'ai exploré quasiment tous les archipels polynésiens (à l'exception de l'archipel des Gambier) :

- archipel de la Société,
- archipel des Australes,
- archipel des Marquises,
- archipel des Tuamotu.

J'ai étudié les différents types de milieux naturels terrestres présents en Polynésie française :

- la cocoteraie des Tuamotu dans le cadre de la valorisation des ressources naturelles polynésiennes,
- la forêt de nuages de l'île de Tahiti dans une optique d'information/préservation,
- les milieux fortement perturbés avec la détection des niches écologiques de l'espèce envahissante *Miconia Calvescens*, ou la détection de certaines espèces rares,
- le milieu agricole avec la cartographie des champs d'ananas de Moorea,
- l'occupation des sols d'une île haute avec la cartographie de Tahiti pour intégration dans un modèle d'érosion,
- le milieu naturel complet d'une île haute avec la cartographie de Moorea.
- Le milieu corallien

J'ai travaillé sur un maximum de données satellitaires ayant un intérêt potentiel pour notre thématique :

- de la donnée optique : Landsat, Aster, Modis-ASTER, Spot, Spot5, Ikonos, Quickbird, WorldView2, ...
- de la donnée radar : JPL-AirSAR, TerraSAR-X.
- des données ancillaires, qu'elles soient géomorphologiques ou météorologiques.
- des données SIG

Ces données de caractéristiques hétérogènes possèdent des résolutions spatiales diverses :

- décamétriques,
- métriques
- sub-métriques.

3. Positionnement scientifique et originalité

Lorsque l'on analyse l'ensemble de ce travail de recherche, on peut constater qu'il se situe à la frontière entre la recherche appliquée et la recherche fondamentale. Je confronte en effet constamment des techniques issues de la recherche fondamentale à des problématiques de recherche appliquée.

En ce sens, mon approche du travail de recherche se situe à l'inverse de ce qui est habituellement la norme dans la communauté du traitement d'images.

En effet, les études sur la végétation présentent souvent un travail partant d'un aspect mathématique ou théorique, puis élaborent un algorithme avec certaines propriétés, et confrontent cet algorithme aux autres algorithmes équivalents sur un site d'étude soigneusement choisi pour éviter au maximum les biais et mettre en avant les propriétés de cet algorithme.

Ce choix évite ainsi la plupart des cas problématiques comme une topographie escarpée, une complexité ou hétérogénéité trop importante, un milieu difficile d'accès, ...

J'ai, pour ma part, adopté une démarche inverse, non pas du fondamental vers l'appliqué mais de l'appliqué vers le fondamental en partant toujours d'une problématique environnementale ou sociétale, et en appliquant des techniques issues de la recherche fondamentale.

Ma préoccupation est souvent moins de créer un nouvel algorithme que de résoudre une problématique environnementale. Si l'on se réfère à l'analogie artistique du préambule, mon but principal est plus d'effectuer un portrait réussi que d'inventer un nouveau type de pinceau. Même si, pour ce faire, je peux être amené à proposer des améliorations sur les outils que je choisis et utilise (cf. l'algorithme *SELF*).

La résolution d'une problématique environnementale place également mon travail de recherche à la frontière entre le traitement d'images et l'écologie. J'en veux pour preuve mes publications, que l'on trouvera aussi bien dans des revues de traitement d'images que des revues d'écologie.

Ce faisant, je contribue à la protection de l'environnement, plaçant mon travail de recherche dans le cadre de la charte française de l'environnement du 1^{er} mars 2005 (cf. *article 6 : La recherche et l'innovation doivent apporter leur concours à la préservation et à la mise en valeur de l'environnement*).

En fin de compte, c'est tout cela qui constitue l'originalité de ce travail : travailler sur un milieu très particulier avec des contraintes fortes, se positionner à la frontière entre l'appliqué et le fondamental en appréhendant ma recherche de la problématique vers la technique, et se situer à la limite entre le traitement d'images et l'écologie.

Pour toutes ces raisons, je suis enclin à penser que mon profil pourrait se rapprocher d'un profil de cogniticien, adoptant une démarche qui consiste à partir de l'analyse des besoins, à mettre en œuvre des techniques diverses en vue de résoudre une problématique. Autrement dit, davantage un profil de compétences qu'un métier.

B. PERSPECTIVES

1. Perspectives en télédétection

Voici ce qui doit être une des premières images satellites de la Polynésie française : une image prise par le satellite espion américain Keyhole 7 (KH-7) à l'époque des premiers essais nucléaires sur l'atoll de Mururoa en 1967 (*figure 62*).

Cette série de satellites (KH-1 lancé en 1963 à KH-38 lancé en 1967) effectuait des photographies de résolution de 1m20 (puis 60cm). Les films argentiques étaient renvoyés sur terre. La mission Keyhole a effectué en tout 19 000 photographies classifiées top secret.

Depuis cette époque l'imagerie satellitaire a fait d'énormes progrès, avec l'avènement des capteurs numériques, multi-spectraux et hyper-spectraux, le radar à synthèse d'ouverture, les progrès de la communication entre le satellite et la terre. On peut parler ici de révolution technologique, tant et si bien que l'imagerie satellitaire est désormais à la base de plus en plus d'applications dans le domaine civil.

Alliée aux techniques de télédétection, de traitement d'images et à la progression constante des moyens informatiques, l'imagerie satellitaire est devenue un outil incontournable pour l'observation de la terre et de ses écosystèmes.

En dépit de progrès indéniables qui sont survenus dans tous les domaines techniques au cours des 40 dernières années, nous n'avons finalement pas progressé autant qu'on aurait pu l'imaginer.

L'outil de façon générale a certes été amélioré, mais la résolution spatiale n'est en fin de comptes pas si différente, les capacités à détecter des objets équivalentes, et surtout la télédétection reste le domaine des spécialistes (militaires et chercheurs).

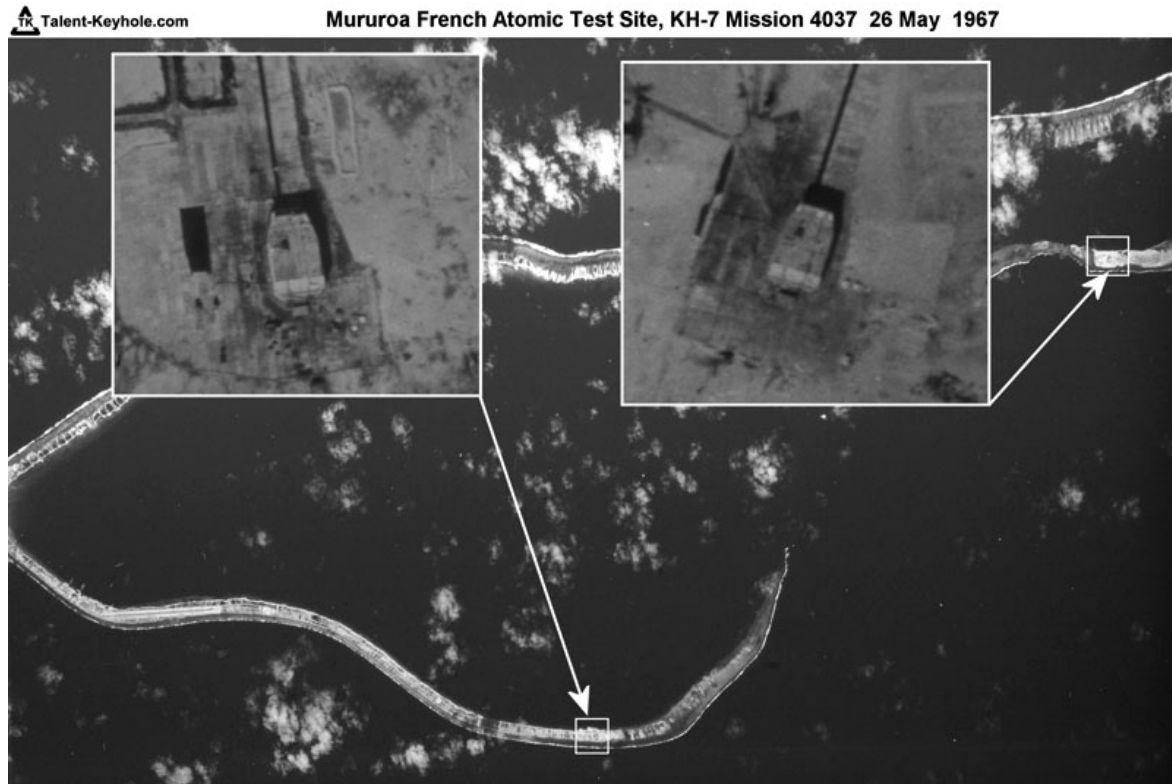


figure 62 Image Keyhole7 de l'atoll de Mururoa (1967)

a. Point d'inflexion

Force est de constater que la télédétection en est arrivée à un point d'inflexion :

La courbe de l'amélioration des capteurs satellites commence se stabiliser. Il suffit de voir le nombre de capteurs à très haute résolution (*THR*) optiques possédant quasiment les mêmes caractéristiques : GeoEye, Quickbird et Pleiade avec des précisions de 50cm en panchromatique et 2m en couleurs (dans cette famille de capteurs, Worldview2 avec ses 8 canaux reste une exception).

Pour ce qui est de l'imagerie optique, le progrès résidera dans la disponibilité de capteurs d'origine militaires avec des précisions de l'ordre de la dizaine de centimètres pour des applications civiles et dans la mise au point de capteurs optiques géostationnaires à haute résolution qui permettront d'établir une surveillance continue d'une région.

Quant à l'imagerie radar, à peu près toutes les configurations de fréquences/polarisations ont été intégrées et testées sur des satellites radar, donnant lieu à un grand nombre d'études sur l'intérêt de tel ou tel capteur. Il reste peu de latitude dans cette direction, si ce n'est qu'en termes de précision, un saut technologique important est en passe d'être réussi avec l'arrivée de capteurs radar à haute résolution (citons TerraSAR-X avec sa précision de 1m).

La technologie arrive aussi à un point d'inflexion en ce sens que les capacités de calculs, la performance des algorithmes font que même les problèmes les plus consommateurs de

ressources informatiques peuvent être menés à bien. Citons le problème du traitement des données hyper-spectrales, les prétraitements automatiques, la surveillance automatique, etc.

b. Démocratisation

Cette synergie entre les capteurs et les capacités de traitements fait que le véritable enjeu s'est déplacé vers une démocratisation de la télédétection : nous sommes encore dans un paradoxe entre le discours des télédéTECTEURS qui font état de potentialités et l'incapacité des utilisateurs à les mener à bien.

Cette démocratisation doit passer par trois phases :

Rendre la donnée accessible : le modèle économique autour de l'image satellite, historiquement basé sur celui de la télécommunication, n'est plus adapté. L'usage des données satellites, encore trop onéreuses, est l'apanage de quelques rares privilégiés en mesure de rentabiliser un tel achat. La donnée devra être obtenue via des financements publics à des tarifs abordables. C'est l'esprit du projet Equipex/Geosud qui finance des images satellites pour les rendre accessibles gratuitement à la communauté scientifique et organismes publics.

Développer les outils informatiques : la télédétection ne doit plus être uniquement réservée aux experts, les outils devront à l'avenir se démocratiser, et devenir utilisables par des néophytes. Par ailleurs, les logiciels propriétaires (souvent considérés comme prohibitifs) devront être remplacés par des logiciels libres (tels que Orpheo Tool Box du CNES, et le projet Quantum Gis). Outre leur gratuité, ces derniers ont l'avantage de s'appuyer sur une communauté d'utilisateurs et d'être en perpétuelle évolution.

Démocratiser les usages : on doit passer d'une économie basée sur le satellite et ses images à une économie basée sur l'utilisateur. L'attention ne se focalisera plus sur la donnée mais sur le produit. En d'autres termes, l'intérêt n'est pas l'image satellite elle-même mais l'application qui en découlera, à l'instar de Google Earth et Google Map qui sont des premiers pas intéressants dans ce sens.

c. Formation

Démocratiser la télédétection et l'utilisation de l'information géographique en général devra passer par un volet formation à grande échelle.

La présence de l'information géographique s'accroît dans tous les domaines de la société. De la même façon que l'on a créé des certifications de compétences transversales en anglais (TOFL), en français (passeport Voltaire) ou en informatique et internet (B2i et C2i), nous pourrions envisager la création d'un certificat de compétences en sciences de l'information géographique. Le programme pourrait regrouper des connaissances théoriques sur les différents types d'informations géographiques (images vecteurs), de projections géographiques, la technique *GPS* grand public, la maîtrise d'un *SIG*, ...

De façon plus pragmatique et à plus court terme, j'envisage de proposer deux formations qui me permettraient d'enseigner la matière dans laquelle je conduis ma recherche et dont je suis le spécialiste à l'UPF, comme c'est la vocation première de tout enseignant-chercheur.

La première concerne le réseau *GPS* et l'utilisation de techniques *GPS* avancées, et la seconde concernerait plus la télédétection et le traitement d'images satellites.

Ces deux cours pourraient compléter l'offre du cours intitulé : « Les outils de la géographie : *SIG* » de la licence de géographie en proposant une partie plus orientée « Télédétection/Images satellites » et/ou « *GPS* ».

Ces cours pourraient aussi être proposés comme un module de master (Master Environnement Insulaire Océanien de l'UPF par exemple), ou un module de thèse. En effet, nombreux sont les étudiants de master ou de thèse à utiliser un *GPS*, des *SIG* ou des techniques basées sur l'information géographique pour leurs travaux de recherche.

Enfin, je pourrais prendre la responsabilité d'animer des cours en lignes à l'UPF, en partenariat avec l'ENSG de l'université de Marnes-La-Vallée, dans le cadre d'une formation diplômante en géomatique.

La géomatique regroupe l'ensemble des sciences et technologies qui permettent de modéliser, de représenter et d'analyser le territoire grâce à l'acquisition et à l'exploitation de données géographiques pour en faire des représentations virtuelles : géolocalisation, imagerie spatiale, bases de données, systèmes d'informations et SIG, technologies du web²³...

Le format des cours en ligne me semble plus adapté car même si le besoin existe, l'étroitesse du marché ne permettra sans doute pas de justifier une filière complète sur ce thème.

Pour information, en septembre 2013, la géomatique a été intégrée dans la liste des « dix professions d'avenir » présentée par le gouvernement français.

2. Perspectives sociétales

a. Atouts et menaces

La Polynésie possède un environnement unique au monde constitué d'un vaste milieu maritime dont la majeure partie est océanique (5 millions de km²) et une partie lagonnaire, et d'îles coralliennes magnifiques dont certaines sont très bien préservées car isolées.

Pour les îles les plus habitées, même si l'expansion de la zone urbaine pose des problématiques de gestion raisonnée, l'intérieur des terres est encore relativement préservé car en grande partie inaccessible et inexploité.

Cet environnement constitue l'atout principal de la Polynésie, mais il est soumis à deux types de pressions :

D'une part les pressions distales telles que le réchauffement planétaire et les changements climatiques, l'acidification des océans, la montée des océans, les pollutions océaniques diverses, etc. Située au centre du Pacifique sud, la Polynésie est pourtant relativement isolée de ses pays frontaliers et subit peu de pressions directes de la part de ceux-ci.

D'autre part les pressions proximales, essentiellement anthropiques, découlant de l'accroissement de la population et de l'expansion de l'urbanisation.

Même s'il est difficile à l'échelle d'un petit pays ultra-marin d'influer sur les pressions distales, une bonne gestion des pressions proximales permet d'en minimiser l'impact.

b. Secteurs prioritaires

Les ressources propres les plus évidentes de la Polynésie sont le tourisme et les ressources halieutiques. D'autres secteurs vont émerger avec la nécessité de voir s'améliorer l'autosuffisance de la Polynésie française, qu'elle soit énergétique avec les énergies

²³ Plaquette : Découvrez le monde de la géomatique, L'école de la géomatique, ENSG, www.ensg.eu

renouvelables (solaire, éolien, ..) diminuant ainsi la dépendance au pétrole, ou alimentaire avec l'agriculture, l'élevage et la pêche.

La montée en puissance de tous ces secteurs requiert une prise de conscience de l'intérêt pour la préservation de l'environnement terrestre, lagonnaire et océanique, de la part de l'administration territoriale mais aussi de la population.

Cette prise de conscience est un pré-requis pour la mise en place d'une gestion raisonnée des ressources naturelles dans un esprit de développement durable.

La Polynésie a les moyens de choisir et maîtriser son propre destin, elle possède un cadre politique d'autonomie lui permettant de légiférer sur les sujets qui la concernent. Elle est donc en mesure de faire des choix politiques forts et de trouver le bon compromis entre le développement de son économie, avec en particulier l'exploitation de ses ressources naturelles, et la préservation de l'environnement à des fins de développement durable mais aussi touristiques.

c. Enjeux de la géomatique et de la télédétection

Pour cela, la géomatique (qui englobe la télédétection) est une science parfaitement adaptée car elle est un outil d'aide à la décision. Elle trouve ses applications dans les domaines de la modélisation et l'aménagement du territoire, la gestion de l'occupation du sol, la maîtrise de l'expansion de la zone urbaine, l'environnement, le développement durable, etc.

Les services territoriaux ont déjà commencé à intégrer la géomatique avec des outils comme la base de données du cadastre, les plans de prévention des risques et les plans généraux d'aménagement des communes. Le site internet Te Fenua²⁴ montre l'intérêt des services territoriaux pour la mise à disposition du public de couches SIG locales.

Dans ce cadre, la télédétection est un outil idéal permettant de produire des couches SIG par la synthèse d'informations géographiques de tous types.

A l'heure actuelle, les deux grands chantiers environnementaux d'importance sont la demande de classement des Marquises au patrimoine mondial de l'UNESCO et le conservatoire maritime de la presqu'île de Tahiti. Mais les grands enjeux de la télédétection sont nombreux et d'origine diverses :

- Gestion durable de l'environnement :

La télédétection peut non seulement aider à établir un état des lieux de l'environnement en fournissant des cartographies thématiques mais aussi permettre la surveillance de cet environnement par des études comparatives sur des zones d'intérêts prioritaires à l'aide de techniques de détection de changements.

Elle procurera des informations essentielles pour contribuer à la modélisation de milieux naturels sensibles. Cette modélisation permettra une gestion environnementale efficace en procurant des informations sur les actions prioritaires et la mesure de l'efficacité des politiques publiques mises en œuvre.

La télédétection pourra aussi aider à mieux appréhender la dynamique environnementale, c'est-à-dire l'évolution en temps réel de l'environnement en fonction d'un certain nombre de facteurs anthropiques ou naturels et ainsi contribuer à préserver les espèces et les écosystèmes.

²⁴ <https://www.tefenua.gov.pf>

- Etudes d'impact sur l'environnement :

Le gouvernement de la Polynésie française et les services territoriaux mettent en place une législation adaptée pour la gestion durable de l'environnement.

La réglementation, comme en France, impose d'effectuer une étude d'impact sur l'environnement lorsqu'un projet dépasse une certaine ampleur. Avec l'intensification de l'urbanisation et l'aménagement du territoire, le nombre d'études va augmenter ainsi que le besoin en personnels qualifiés en géomatique. Les SIG sont en effet devenu un outil incontournable pour ce type d'études. Les données issues de la télédétection seront de plus en plus utilisées dans ce cadre.

- Prévention des risques :

La télédétection a son rôle à jouer lors de catastrophes naturelles comme les cyclones, tsunamis, glissements de terrain, montée globale des eaux, ou autres. Un besoin de quantification rapide des dégâts va émerger ne serait ce qu'à des fins d'indemnisations par les assurances.

Il faut savoir qu'une charte catastrophe internationale a été signée par la plupart des compagnies fournissant de la donnée satellite. Cette charte engage les compagnies signataires à fournir prioritairement et gratuitement des données satellites sur les lieux de catastrophes. Le gouvernement de la Polynésie française ou les services du haut-commissariat en Polynésie française qui est chargé de la sécurité civile sont habilités à faire appel à ces prérogatives.

Enfin, la télédétection alliée à la géomatique peut aussi être utilisée pour la gestion des risques sanitaires par la modélisation de la propagation de maladies liées aux moustiques comme la Dengue, le Chikungunya, mais aussi de maladies dues à la mauvaise santé du récif corallien comme la gratte, etc.

- Surveillance :

La Polynésie doit acquérir la maîtrise de son espace maritime, ce qu'elle n'a pas les moyens de faire efficacement pour le moment. L'utilisation d'images satellites permettra une surveillance de la ZEE de la Polynésie française, en particulier pour le contrôle du trafic maritime, les pollutions éventuelles, mais surtout pour lutter contre la pêche illicite.

d. Aide à la décision

Le territoire a déjà commencé à se doter de moyens SIG pour un certain nombre d'applications comme le cadastre, les plans publics d'aménagement, les zones d'aires marines protégées. Il faut continuer et intensifier cet effort.

En effet, il y a un besoin de cartographie globale à l'échelle de la Polynésie avec entre autres un carte d'occupation des sols. En métropole, une cartographie globale Corine Landcover²⁵ au pas de 20 mètres est disponible selon une nomenclature normalisée. Ce type de couverture devra être produit pour l'ensemble des îles polynésiennes. Pour ma part, je préconiserais une définition spatiale plus fine, en adéquation avec la petite taille des îles et qui serait tout à fait envisageable financièrement au vu de la faible surface cumulée des îles. Cela pourrait être réalisé à partir d'images métriques (Spot5 ou équivalent).

Les problématiques importantes sont l'occupation du sol ainsi que le bâti en 3 dimensions, ce qui inclue la hauteur de la végétation (forêt + arbres isolés). Cela devrait être disponible a

²⁵ <http://www.statistiques.developpement-durable.gouv.fr/donnees-ligne/li/1825/1097/occupation-sols-corine-land-cover.html>

minima sur les zones urbaines des îles les plus peuplées. Ces données serviraient de base à la mise au point de projets d'aménagement.

L'originalité de la Polynésie c'est sa vaste étendue lagonnaire. Cet espace nécessite lui aussi un grand besoin de gestion raisonnée. Des informations sur le lagon et le littoral devraient être incluses dans les SIG territoriaux. Une cartographie comportant le littoral avec son bâti, la nature de ses fonds (sable blanc, sable noir, vase, etc.) mais aussi le chenal avec la bathymétrie, le récif corallien (sur un jeu de classes réduit), les zones à fort courant, la signalisation maritime, les passes etc. Bref, toutes les informations permettant une utilisation raisonnée du lagon.

Ces informations constitueront les données de base pour un SIG territorial ambitieux qui deviendra ainsi un véritable outil pour les gestionnaires.

e. Caractéristiques des images

Un projet de construction d'une antenne de réception satellite en Polynésie française est en cours de préparation, avec pour principale vocation le contrôle de l'espace maritime. Cela étant, si ce projet aboutissait, des besoins pour traiter cette donnée dans tous les domaines précédemment cités seraient sans doute exprimés, de sorte que les acteurs locaux devraient faciliter l'accès à cette donnée à d'autres fins.

Soulignons tout de même que les données satellites devront avoir des caractéristiques adaptées aux spécificités de la Polynésie dans les différentes problématiques.

La surveillance de l'espace maritime se fera à l'aide d'images radar avec une définition décimétrique.

Les images optiques ne pourront être envisagées que pour des études nécessitant une périodicité annuelle au plus, la couverture nuageuse rendant incertaines les images exploitables. Un moyen de contourner ce problème consiste à traiter l'image au niveau du pixel tout en gardant uniquement les pixels exploitables de chaque image. Pour des images à une seule date avec des périodicités plus courtes, les images radar seront choisies.

La surveillance de l'environnement devra utiliser conjointement des images optiques, radar et autres métriques ou centimétriques selon la problématique.

Les images optiques devront obligatoirement inclure la totalité des bandes disponibles et en particulier l'infrarouge pour les applications sur la végétation.

Enfin, un effort devra être fait pour améliorer la couverture en données météorologiques qui gagnerait à être plus homogène sur les îles principales, car le manque de stations météorologiques dans l'intérieur des terres est un frein aux études environnementales.

3. Perspectives de recherche

Ce mémoire d'habilitation à diriger des recherches n'étant bien sûr qu'un point d'étape dans mon parcours de recherche, certains aspects du travail présenté ici sont encore en chantier comme l'étude des techniques de fouille de données sur la cocoteraie des Tuamotu, l'étude de l'apport de la texture optique ou radar pour la classification du couvert végétal polynésien, l'analyse et l'optimisation de certains facteurs abiotiques comme l'insolation, les précipitations, la modélisation de facteurs environnementaux, etc.

Je continuerai naturellement à travailler sur ces aspects au fur et à mesure de mon parcours. Parallèlement, les nouvelles perspectives ne manquent pas.

La télédétection environnementale est un domaine en plein essor et mes perspectives de recherche émaneront de plusieurs domaines : les progrès de l'imagerie satellitaire avec le lancement de nouveaux capteurs, les nouveaux algorithmes et de nouvelles techniques qui découleront des progrès des techniques de traitement d'images, mais aussi les synergies créées entre des sciences jusque-là indépendantes.

a. Capteurs

Le lancement régulier de nouveaux capteurs satellitaires (leur durée de vie dépasse rarement une dizaine d'années) va nous offrir l'opportunité d'étudier de nouvelles données avec des caractéristiques toujours plus intéressantes, que ce soit au niveau spatial et spectral mais aussi au niveau de la répétitivité et de la rapidité de programmation.

Je vais maintenir ma veille technologique sur les nouveaux capteurs afin d'en étudier les capacités sur les problématiques polynésiennes.

Mentionnons ici quelques capteurs optiques ou radars récents à très haute résolution sur lesquels je n'ai pas encore eu l'opportunité de travailler, ou qui doivent être lancés à courte échéance :

- GeoEye1 et GeoEye2 (~40cm en panchromatique et ~1.5m en R, V, B, Proche infrarouge)
- Worldview2 (~40cm en panchromatique et ~1.8m sur 8 couleurs)
- Spot6 et Spot7 (~1.35m en panchromatique et ~6m en R, V, B, Proche infrarouge)
- Pléiade (~50cm en panchromatique et ~2m en R, V, B, Proche infrarouge)
- Tandem-X (bande X, ~1m/3m/18m selon les modes de polarisation)
- Radarsat2 (bande C, 3m/10m en polarisation HH-HV-VH-VV)

Précisons que, comme démontré dans ce document, l'avenir se trouve dans la combinaison de plusieurs capteurs aux caractéristiques complémentaires, comme optique, radar et ancillaires mais aussi en combinant l'imagerie très haute résolution et moyenne résolution.

J'ai obtenu d'excellents résultats en combinant plusieurs types d'informations sur la végétation : spectrales (imagerie optique), structurelles (imagerie radar), texturales (textures radar et optique), et contextuelles (données ancillaires).

Je projette d'étudier l'apport d'un autre type de capteur : le capteur Lidar permettant d'avoir accès à un nouveau type d'information qui est la structure verticale des couverts forestiers. En effet, le capteur Lidar est un capteur actif qui émet un signal laser, et enregistre la réponse du milieu, celle-ci va renvoyer un signal sur le sol, mais aussi sur les différents étages de la végétation, du sous-bois à la canopée.

Je vais aussi m'intéresser aux capteurs larges champs sur des applications plus océaniques (SMOS et AQUARIUS capteurs de salinité de l'océan et d'humidité des sols), aux capteurs hyper-spectraux hautes résolution (HYPXIM, Hyperion, CHRIS/PROBA), ainsi qu'aux capteurs spécifiques pour la végétation (PROBA-V capteur permettant d'obtenir un indice de végétation à large échelle) ou la biomasse (BIOMASS capteur radar).

b. Techniques

Les techniques de traitement d'image sont aussi en plein essor puisque l'information géographique et les images satellites prennent une part de plus en plus importante dans notre société. Il est évidemment difficile de prédire quelles techniques vont s'imposer dans les années à venir, mais je voudrais évoquer deux pistes qui me semblent prometteuses.

La première est l'intégration d'informations *a priori* sur la distribution des classes comme contrainte dans nos algorithmes de classification. A titre d'exemple, la forêt de nuages n'est présente qu'au-dessus d'une certaine altitude, qui varie selon l'orientation de la côte considérée. Des techniques, inspirées de celles mises en œuvre dans des systèmes experts, pourraient être une façon d'aborder le problème en permettant l'intégration d'un certain nombre de règles quantitatives et qualitatives issues des connaissances des biologistes sur les espèces végétales.

La seconde est la technique récente de *SVM* multi-noyaux (ou multi-Kernel). Nous avons vu que pour classifier des données hétérogènes, il est plus intéressant de mettre en œuvre une technique de fusion *SVM* qui consiste à classifier des résultats de classifications mono-sources. Cette technique a pour avantage d'exploiter le schéma d'erreur individuel des classes pour améliorer les résultats, mais elle souffre du fait qu'il faut optimiser $N+1$ algorithmes *SVM* si l'on a N types de données différentes.

La technique des *SVM* multi-noyaux permet de classifier N types de données avec N noyaux différents en une seule fois ; il est intéressant de comparer cette technique naissante avec la fusion *SVM* classique.

De façon plus générale, je vais m'atteler à améliorer la généralité de mes méthodologies de télédétection. L'idée est de faire en sorte qu'une méthode ou une chaîne de traitement soit exportable sur un autre site d'étude, une autre île. Cela permettra d'optimiser la télédétection. Ce point est encore compliqué à mettre en œuvre car les images ont dès le départ des caractéristiques de prises de vues différentes et les outils de traitement d'images voient ici leurs limites.

J'ai été confronté à cette problématique lors de l'étude sur la classification de l'ensemble des cocoteraies des Tuamotu sur des images Ikonos. Des techniques de fouilles de données ont permis de faire un premier pas vers cette généralité, mais ce point doit faire l'objet d'études complémentaires.

Enfin, un dernier aspect technique que j'aimerais explorer, est le fait de fusionner l'information plutôt que la donnée. En effet, lorsque l'on possède plusieurs acquisitions d'une région, cela permet de tirer partie de tous les pixels utiles de chaque image, permettant ainsi de pallier le problème des nuages, récurrent sur les îles tropicales.

c. Thématiques

Une ouverture vers d'autres thématiques me permettra aussi d'élargir mes perspectives de recherches ; citons quelques exemples :

Le milieu corallien pourrait être abordé avec des images Worldview2 dont les caractéristiques sont particulièrement adaptées au milieu marin (50 cm de résolution spatiale et 8 bandes spectrales dont 4 spécialement adaptées au milieu marin).

La télédétection océanique avec des capteurs à plus basse résolution permettant de détecter la salinité de la mer, le plancton, l'état de surface, les courants, etc.

L'utilisation de ces techniques océaniques en milieu lagonnaire, pour modéliser et surveiller les facteurs biologiques favorables à la culture des huîtres perlières. La taille de la plupart des atolls (une centaine de km²) imposera l'utilisation de capteurs haute résolution ou très haute résolution.

Les sciences de la vie et de la terre ont un besoin croissant d'informations géographiques pouvant être obtenues par des techniques de télédétection.

Les domaines pouvant être abordés par des techniques de télédétection sont pléthores, et la difficulté réside essentiellement dans le fait de ne pas se disperser. Une étude doit être envisagée uniquement car elle apporte une potentialité en termes de recherche et de publications et non simplement pour fournir un service.

d. SIG versus télédétection

Historiquement, la télédétection et les *SIG* ont suivi un parcours parallèle mais l'interaction entre ces deux domaines a toujours été faible.

Les *SIG* peuvent être considérés comme les descendants de la photogrammétrie, et sont basés sur les vecteurs alors que la télédétection vient de la mesure d'une valeur physique, elle a évolué en regroupant plusieurs valeurs de mesures (pixels) en une image (raster).

Ces deux communautés ne parlent pas le même langage et ne se comprennent que très peu. Il s'agit en fait de deux approches différentes du traitement de la même information, l'information géographique. Cela revient à comparer le dessin (trait - vecteur) à la peinture (tâches de couleurs - pixel).

L'avenir viendra de la fusion de ces deux sciences et c'est sur l'écotone entre les *SIG* et la télédétection que l'information géographique se développera.

L'étude sur la modélisation de la santé du récif frangeant de l'île de Moorea précédemment développée peut être considérée comme une première approche de cette fusion. Un indice de santé du récif corallien a été modélisé à partir de données de type *SIG* et de données issues de la télédétection, permettant d'avoir une idée de l'importance relative de certains facteurs hétérogènes sur la santé du corail et donc d'optimiser les campagnes de préservation de ce dernier. Ce type d'études devra être encouragé car elles permettent de faire le lien entre la science et les décideurs publics en leur donnant des outils de compréhension et de gestion de l'environnement.

e. Études exhaustives versus télédétection

De la même façon, il y a toujours eu une méfiance réciproque entre les écologues qui ont une connaissance parfaite du terrain et les télédéTECTEURS. Les uns jetant un regard souvent critique sur la qualité des cartographies satellitaires et les potentialités de la télédétection, et les autres essayant d'ignorer les questions dérangementantes des spécialistes des milieux naturels.

Pour expliquer mon point de vue sur ce sujet, il faut revenir à l'analogie entre la télédétection et la peinture : avant l'avènement des techniques photographiques, la peinture était essentiellement figurative. Son but était de faire une représentation aussi fidèle que possible du sujet. Mais à l'arrivée de la photographie, ce rôle est devenu caduc et a muté vers une expression plus artistique, et a tenté de représenter les impressions, sentiments, et émotions que la photographie avait plus de mal à restituer.

De la même façon que la photographie a entraîné une évolution de la peinture vers des aspects plus conceptuels et moins représentatifs, la qualité et la précision du travail des écologues induisent que l'avenir de la cartographie n'est pas dans la représentation exacte d'une situation mais dans la représentation synthétique d'une réalité thématique. La comparaison avec une étude terrain exhaustive n'a donc pas de sens. En revanche, confronter ces deux approches sera riche d'apprentissages de part et d'autres, et ouvrira de nombreuses perspectives de recherches.

f. Sujets de recherche

Il est toujours délicat de se projeter sur ce que l'on proposera comme sujet de recherche à moyen et long terme, tant les techniques évoluent vite et les sources de financements aléatoires. De plus, beaucoup d'études trouvent leur origine dans des opportunités qu'il est difficile de prévoir.

Le début de ce chapitre liste les principales perspectives qu'il me semble intéressant de développer. Certaines pourront faire l'objet de stages de master, ou de licence et d'autres d'un sujet de thèse.

Les difficultés dans le financement et l'acquisition d'images font que lorsque je propose un sujet de thèse, je choisis toujours un site d'étude et une problématique sur lesquelles j'ai déjà suffisamment d'éléments pour commencer à travailler. Un premier jeu d'images satellites, et un bon contexte avec une problématique porteuse sont un minimum pour avoir une base solide.

Ainsi, travailler sur l'île de Raiatea pourrait être un sujet de thèse. Cette île est la seconde en termes de population en Polynésie française et c'est le lieu de nombreux enjeux environnementaux par la conjonction d'une population en expansion, et d'un milieu naturel relativement préservé avec une forêt de nuage présentant un fort taux d'endémisme mais avec des menaces identifiées comme l'invasion du *Miconia*, ou la disparition progressive du *Tiare apetahi*. Le domaine territorial de Faaroa pourrait être un site d'étude intéressant.

S'agissant des images, nous possédons déjà de la donnée AirSAR de 2000, celle-ci pourrait être complétée par de la donnée Tandem-X, et une image Worldview2 (*figure 63*) dont les propriétés spectrales (8 bandes) et spatiales (60cm) n'ont jamais été étudiées dans le cadre polynésien. De la donnée Lidar ainsi que des données ancillaires météorologiques ou tirées d'un MNE permettront de rajouter des informations complémentaires.

Dans ce cadre, plusieurs pistes sont envisagées : étude des textures radar et optique à très haute résolutions, comparaison d'un schéma de fusion SVM classique avec un algorithme SVM multi-noyaux, classification sous contrainte, espèces rares et espèces envahissantes, modélisation environnementales, cartographie lagonnaire, etc. De surcroît, l'île de Raiatea et le domaine de Faaroa en particulier ont fait l'objet depuis quelques années d'études terrain intensive de la part de la direction à l'environnement, ces études pourraient nous permettre de confronter les études exhaustives avec les études de télédétection.



figure 63 Image Worlview2 de Raiatea et Tahaa (2011)

CONCLUSION

La rédaction d'un mémoire d'HDR est une étape importante dans la carrière d'un enseignant-chercheur : celui-ci a pour but d'obtenir l'habilitation à diriger des recherches, c'est-à-dire que l'on considère que l'enseignant-chercheur a acquis l'envergure nécessaire pour pouvoir conduire ses recherches de manière autonome.

Concrètement, cela signifie qu'il peut dorénavant proposer lui-même ses axes de recherches, chercher des financements, encadrer des thésards. Au cours du cheminement que fut l'élaboration de ce mémoire, j'ai pris conscience que cette autonomie caractérisait déjà fortement mon parcours de chercheur.

Cependant, au moment de conclure, loin de considérer ce travail comme un aparté administratif, je prends pleinement la mesure de tout ce que cette démarche m'a apporté de positif, et qui fait de cette préparation d'HDR une étape clé de ma carrière.

Ce fut une sorte de pause réflexive et structurante au cours de laquelle j'ai pris le temps d'inventorier l'intégralité des études que j'ai effectuées depuis mon recrutement à l'université de la Polynésie française, d'organiser ce travail et de le mettre en perspective.

Ce travail de synthèse m'a non seulement permis de prendre du recul sur le travail réalisé, mais il m'a aussi fait prendre conscience de la singularité de mon profil : j'ai en effet recueilli des avis extérieurs au travers des discussions que j'ai pu avoir avec mon directeur d'HDR et quelques collègues qui m'ont appuyé dans cette démarche.

J'ai été conforté dans mes choix de carrière, convaincu que la télédétection est une thématique devenue incontournable, en particulier dans ce territoire ultra-marin présentant des enjeux importants et des pistes foisonnantes.

Ce temps de réflexion m'a ouvert de nouveaux horizons en renforçant mon intuition selon laquelle de nouveaux types de collaborations inédites créeraient des synergies prometteuses. Ce fut alors l'occasion de faire le point sur l'éventail des possibilités qui s'offrent à moi.

Devant tant d'opportunités, l'écueil à éviter c'est l'éparpillement. Fort de ce nouvel éclairage porté sur mon travail et mon profil et gardant bien à l'esprit ce contexte spécifique dans lequel j'évolue, j'ai pu me recentrer. Des pistes mieux ciblées, mieux définies, notamment celles susceptibles d'être les plus cohérentes à tous les égards et les plus porteuses me permettront de me projeter plus sereinement dans ce que je souhaite faire par la suite. Ainsi ce document sera un guide pour les années futures.

Ce travail est constitutif de mon parcours, il était nécessaire. Il m'a permis d'acquérir une maturité, d'envisager la suite de ma carrière avec une motivation accrue.

« Les tableaux ne sont que recherches et expériences. » Pablo Picasso

LISTE DES PUBLICATIONS

Articles dans des revues internationales à comité de lecture

ecoinfo2012: R. Pouteau, J.-Y. Meyer, R. Taputuarai et B. Stoll, Support Vector Machines to Map Rare and Endangered Native Plants in Pacific Islands Forests. *Ecological Informatics* **9**: 37-46, 2012.

jstars2012: R. Pouteau et B. Stoll, SVM Selective Fusion (SELF) for Multi-sources Classification of Structurally Complex Tropical Rainforest. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **5**(4): 1203-1212, August 2012.

ecomod2011: R. Pouteau, J.-Y. Meyer et B. Stoll, A SVM-based Model for Predicting Distribution of Invasive Tree *Miconia Calvescens* in Tropical Rainforests. *Ecological Modelling* **222**: p. 2631-2641, 2011.

grsl2010: C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, B. Stoll, B. Fruneau et J.-P. Rudant, Classification of Tropical Vegetation Using Multifrequency Partial SAR Polarimetry. *IEEE Geoscience and Remote Sensing Letters* **8**(1): p. 133-137, jan. 2010.

tgrs2009: C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, B. Stoll, B. Fruneau et J.-P. Rudant, Support Vector Machine for Multi-frequency SAR Polarimetric Data Classification. *IEEE Transaction on Geosciences and Remote Sensing* **47**(12): p. 4143-4152, dec. 2009.

Articles dans des revues électroniques à comité de lecture

vertigo2010: R. Pouteau, J.-Y. Meyer, R. Taputuarai et B. Stoll, La fonte de la biodiversité dans les îles : modélisation de l'impact du réchauffement global sur la végétation orophile de Tahiti (Polynésie française). *Vertigo - la revue électronique en sciences de l'environnement [En ligne]* dec. 2010.

Publications dans des conférences internationales à comité de lecture

igarss2013: R. Pouteau, a. Collin, P. Archambault et B. Stoll, Modeling Reef Health From Upstream Socio-Ecological Components Using GIS And RS. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'13*, Melbourne, Australia, July 2013, IEEE.

aprs2012: S. Chabrier, J.-B. Goujeon et B. Stoll, SVM texture classification for tropical vegetation mapping. *SPIE Asia-Pacific Remote Sensing*, Kyoto, Japan, 29 Octobre - 1 Novembre 2012.

icdim2011: E. Desmier, F. Flouvat, B. Stoll et N. Selmaoui-Folcher, Coconut Fields Classification Using Data Mining on a Large Database of High -Resolution Ikonos Images. *Sixth International Conference on Digital Information Management*, Melbourne, Australia, 14-16 Sept. 2011, SPIE.

isrse2011: R. Pouteau, A. Collin et B. Stoll, A Comparison of Machine Learning Algorithms for Classification of Tropical Ecosystems Observed by Multiple Sensors at Multiple Scales. *34th International Symposium on Remote Sensing of Environment*, Sydney 2011, 10-15 avr. 2011.

- isrse2011:** R. Pouteau, C. Lardeux, B. Stoll et S. Chabrier, Toward an Optimal Fusion Scheme for Multisource Vegetation Classification. *34th International Symposium on Remote Sensing of Environment*, Sydney 2011, 10-15 avr. 2011.
- isrse2011:** R. Pouteau, J.-Y. Meyer, R. Taputuarai et B. Stoll, A comparison between GARP model and SVM regression to predict species potential distribution: the case of *Miconia calvescens* on Moorea, French Polynesia. *34th International Symposium on Remote Sensing of Environment*, Sydney 2011, 10-15 avr. 2011.
- igarss2011:** R. Pouteau et B. Stoll, "Fusion for Classes in Difficulty" for Accurate and Speed Tropical Rainforests Classification. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'11*, Vancouver, Canada, 24-29 juil. 2011, IEEE.
- igarss2010a:** R. Pouteau, B. Stoll et S. Chabrier, Multi-sources SVM Fusion for Environmental Monitoring in the Marquesas Archipelago. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10*, Honolulu, Hawaï, USA, p. 2719-2722, 25-30 juil. 2010, IEEE.
- igarss2010b:** R. Pouteau, B. Stoll et S. Chabrier, Ground Truth Method Assessment for SVM Based Landscape Classification. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'10*, Honolulu, Hawaï, USA, p. 2715-2718, 25-30 juil. 2010, IEEE.
- ipta2010:** R. Pouteau, B. Stoll et S. Chabrier, Support Vector Machine Fusion of Multisensor Imagery in Tropical Ecosystems. *International Conference on Image Processing Theory, Tools and Applications, IPTA'10*, Paris, France, p. 325-329, 7-10 juil. 2010, IEEE.
- spie2009:** R. Teina, D. Béréziat, B. Stoll et S. Chabrier, A Comparative Study of Several Supervised Classifiers for Coconut Palm Tree Fields Type Mapping on 80cm RGB Pansharpened Ikonos Images. *Image Processing Machine Vision Applications II, Proceeding of SPIE (SPIE, Bellingham, WA 2009) 72510X*, 7251, 2009.
- igarss2008:** P.-L. Frison, C. Lardeux, J.-C. Souyris, C. Tison, B. Stoll et J.-P. Rudant, Radar Polar Decomposition for Natural Surfaces Cartography *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'08*, Boston, Massachusetts, U.S.A., 4, p. 363-366, 6-11 juil. 2008, IEEE.
- spie2008:** R. Teina, D. Béréziat et B. Stoll, A Spatial Poisson Process to Classify Coconut Fields on Ikonos Pansharpened Images. *SPIE Multispectral, Hyperspectral, and Ultraspectral Remote Sensing Technology, Techniques, and Applications II*, Nouméa, New Caledonia, 7149, (10 pages), 17-21 nov. 2008.
- igarss2008:** R. Teina, D. Béréziat, B. Stoll et S. Chabrier, Toward a Global Tuamotu Archipelago Coconut Trees Sensing using High Resolution Optical Data. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'08*, Boston, Massachusetts, U.S.A., 2, p. 797-800 6-11 juil. 2008, IEEE.
- envisat2007:** C. Lardeux, P.-L. Frison, C. Tison, D. Deleflie, J.-C. Souyris, J.-P. Rudant et B. Stoll, Comparison of the ASAR Alternative Polarisation Mode to Full Polarimetric Acquisition for Land Use Estimation Over Tropical Regions. *ENVISAT'07*, Montreux, Suisse, 23-27 avril 2007.
- polinsar2007:** C. Lardeux, P.-L. Frison, C. Tison, D. Deleflie, J.-C. Souyris, J.-P. Rudant et B. Stoll, Comparison of Compact Polarimetric with Full Polarimetric Radar Data for

Land Use Discrimination Based on SVM Classification. *POLinSAR '07*, Frascati, Italy, 22-26 jan. 2007.

alos2007: C. Lardeux, P.-L. Frison, C. Tison, J.-C. Souyris, S. Niculescu, B. Stoll et J.-P. Rudant, Contribution of ALOS PALSAR data for land use estimation. *1st Joint Pi Symposium, ALOS '07*, Kyoto, Japan, 19-23 nov. 2007.

majecstic2007: R. Teina, D. Béréziat et B. Stoll, Étude de la cocoteraie des Tuamotu sur des images Ikonos. *MAJECSTIC '07*, Caen, Prépublication 38, p. 265-269, 29-31 oct. 2007, Schedae 07, Presses universitaires de Caen.

igarss2006: C. Lardeux, P.-L. Frison, J.-P. Rudant, J.-C. Souyris, C. Tison et B. Stoll, Use of the SVM Classification with Polarimetric SAR Data for Land Use Cartography. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'06*, Denver, Colorado, p. 493-496, 31 juil.- 04 aout 2006, IEEE.

igarss2005: B. Stoll, SARvi: A Vegetation Index Based on AirSAR data for South Pacific Volcanic Islands Vegetation mapping. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'05*, Séoul, Korea, 6, p. 4331-4334, 25-29 juil. 2005, IEEE.

igarss2004: B. Stoll et P. Capolsini, A Simple Class-Set Based Vegetation Classification of a South Pacific Island (Moorea Island, French Polynesia) Using Both AirSAR and MASTER Data. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'04*, Anchorage, Alaska, 4, p. 2688-2691, 20-24 sept. 2004, IEEE.

Présentations sans publications (Posters)

spie2008: C. Lardeux, P.-L. Frison, J.-C. Souyris, C. Tison, B. Stoll et J.-P. Rudant, Contribution of radar polarimetric data for the cartography in tropical environment. *SPIE Asia-Pacific Remote Sensing, Nouméa, New Caledonia, SPIE*, 17-21 nov. 2008.

psi2008: R. Pouteau et B. Stoll, Project on Support Vector Machines (SVM) Fusion for Classification of Multisensor Data: Application to Marquesas Archipelago Vegetation Mapping. *PSI'08, Tahiti - Polynésie française*,

perf2006: B. Stoll et C. Lardeux, Téledétection du Couvert Végétal des Tubuai (Australes) sur Données Radar et Optique à Haute Définition. *Point d'étape de la Recherche dans le Pacifique, Papeete - TAHITI - Polynésie française*,

arfp2004: B. Stoll et P. Capolsini, Classification de la vallée d'Opunohu Valley (Moorea - Polynésie Française) Utilisant des Données AirSAR et MASTER. *Assises de la Recherche Française dans le Pacifique, Nouméa, Nouvelle-Calédonie, 23-27 Août 2004*.

igarss2003: P. Capolsini, B. Stoll et S. Andrefouet, A Comparison of Classification Algorithms for Coral Reef Mapping in South Pacific Island. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS'03*, Toulouse, France, IEEE, 21-25 juil. 2003.

Co-encadrements de thèses

thesis2011: Thèse de Doctorat de R. Pouteau, Etude de la phytogéographie des îles hautes de Polynésie française par classification SVM d'images multi-sources. *Science, Tahiti, Université de la Polynésie française, 07 nov. 2011*.

thesis2009: Thèse de Doctorat de R. Teina, Caractérisation de la Cocoteraie des Tuamotu à Partir d'Images Satellites à Très Haute Résolution Spatiale. Paris, Université Pierre et Marie Curie UPMC, 15 sept. 2009.

thesis2008: Thèse de Doctorat de C. Lardeux, Apport des Données Radar Polarimétriques pour la Cartographie en Milieu Tropical. Science de l'Information Géographique, Marnes-La-Vallée, Université de Paris-Est - Laboratoire GGI - Equipe OTIG, 09 dec. 2008.

Rapports

report2011: Rapport d'étude de J.-B. Goujeon, Optimisation de la classification de donnée multi-spectrale satellitaire THR, application sur les îles hautes Polynésiennes., Anger, groupe ESEO, 13/06/2011 au 09/09/2011.

report2010: Rapport d'étude de R. Ballet, Classification Thématique de l'Occupation des Sols de l'Île de Tahiti sur Données Quickbird et SPOT5. 2010.

report2010: Rapport d'étude de E. Desmier, Détection des cocoteraies sur les images satellites des Tuamotu-Gambier. Nouméa, 2010.

report2009: Rapport d'étude de M. Castela, Pertinence des Attributs de Textures pour la Classification SVM du Milieu Naturel des Îles Hautes Polynésiennes. Application au Plateau de Taravao. Grenoble, INPG ATIS, 15/02/2009 au 28/08/2009.

report2009: Rapport d'étude de R. Pouteau et B. Stoll, Etude de la Cartographie du Milieu Naturel de Moorea par Classification SVM Conjointe de Données Satellites Radar et Optique à Très Haute Résolution. 2009.

report2007: Rapport d'étude de R. Teina, Etude pilote de la cocoteraie des Tuamotu sur des images Ikonos - Tikehau. Novembre 2006 - Octobre 2007.

report2005: Rapport d'étude de C. Lardeux, Apport de la Polarimétrie Radar à la Cartographie Thématique en Polynésie française. Mars - Août 2005.

report2004: Rapport d'étude de R. Provin, Cartographie de la végétation des îles Australes à partir d'images AirSAR ou MASTER. Juin à Août 2004.

report2003: Rapport d'étude de J. Thomas, Etude de la Cartographie du Couvert Végétal des Îles Hautes Polynésiennes. juin à Août 2003.

Télédétection environnementale en Polynésie française

Apport de la classification SVM pour la cartographie du couvert végétal en milieu insulaire tropical

Résumé : J'ai été recruté à l'université de la Polynésie française en 2001 sur un poste de maître de conférences en informatique et télédétection. J'enseigne essentiellement l'informatique (algorithmique, programmation C et technologies web) au sein du département des Sciences. Je suis responsable de la certification informatique et internet (C2i niveau 1 et C2i2e).



La thématique de recherche que je me suis attaché à développer depuis mon arrivée à l'UPF est l'étude de la répartition de la végétation des atolls et îles hautes de Polynésie française par télédétection satellitaire. Le couvert végétal des îles hautes polynésiennes est caractérisé par une forte hétérogénéité et une forte complexité spatiale principalement dues à la topographie escarpée, au phénomène de Foehn et aux microclimats qui en résultent. Cette complexité spatiale est une contrainte forte pour les techniques de traitement d'images couramment utilisées en télédétection et impose l'utilisation de données aux propriétés complémentaires. J'ai conduit mes recherches selon deux axes : l'évaluation de la pertinence de chaque type d'images satellitaires et l'étude comparée des techniques de traitement d'images et en particulier des différents algorithmes de classification. Si l'on ne devait retenir qu'un résultat comme le plus représentatif de l'ensemble de ce travail, ce serait l'optimisation de la classification de la végétation des îles hautes par fusion SVM de données hétérogènes (algorithme SELF).

Cinq articles dans des revues internationales à comité de lecture ont été publiés, ainsi qu'un article dans une revue nationale en ligne à comité de lecture et vingt deux actes de conférences internationales, là-aussi avec comité de lecture. En outre, j'ai co-encadré trois thèses et participé à la rédaction de huit rapports d'études.

Mots clés : télédétection environnementale, classification supervisée, fusion SVM, couvert végétal des îles polynésiennes

Environmental remote sensing in French Polynesia

SVM classification contribution to the vegetation cover cartography in tropical insular environment

Abstract : I was recruited at the University of French Polynesia in 2001 as a computing and remote-sensing assistant professor. I am essentially teaching computer sciences (algorithmics, C language, web technologies) in the sciences department. I am responsible for the computing and internet certification (C2i level1 and C2i2e).

Since my arrival at the UPF, I have been involved in the study of the high islands and atolls vegetation using satellite remote sensing. The French Polynesian high island vegetation cover is characterized by a strong heterogeneity and high spatial complexity, which are caused by the steep topography, the Foehn phenomenon, and the resulting micro-climates. This spatial complexity is a strong constraint on the usual image processing technics commonly used in remote-sensing; it imposes the use of complementary data. I carried out my researches along two axis: the evaluation of the relevance of each type of satellite image data, and the comparative study of the image processing technics, in particular the various classification algorithms. The most representative result of this whole work is the optimization of the high islands vegetation classification using support vector machine fusion of heterogeneous data (SELF algorithm).

Five international peer-reviewed journal articles have been published, along with one national peer-reviewed online article and twenty-two international peer-reviewed conferences proceedings papers. I also co-directed three thesis, and participated to height study reports.

Key words : environmental remote sensing, supervised classification, SVM fusion, Polynesian islands vegetation cover