



HAL
open science

Apport de l'analyse fonctionnelle pour l'estimation de variables environnementales à partir de formes d'onde LiDAR

M. Mezian

► **To cite this version:**

M. Mezian. Apport de l'analyse fonctionnelle pour l'estimation de variables environnementales à partir de formes d'onde LiDAR. Sciences de l'environnement. 2014. hal-02600091

HAL Id: hal-02600091

<https://hal.inrae.fr/hal-02600091>

Submitted on 16 May 2020

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

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



MASTER 2 PHYSIQUE ET INGÉNIERIE
PARCOURS PHYSIQUE INFORMATIQUE 2013-2014

Apport de l'analyse fonctionnelle pour l'estimation de variables environnementales à partir de formes d'onde LiDAR

Rapport de stage de fin d'études

Auteur :
Miloud MEZIAN

Responsables de stage

M. Nicolas BAGHDADI (UMR TETIS, Irstea)
M. Jean-Stéphane BAILLY (UMR TETIS, AgroParisTech)

Responsables de stage à l'UM2

M. David CASSAGNE
M. Hervé DICKY

Table des matières

| | | |
|----------|--|-----------|
| I | La technologie LiDAR : Etat de l'art | 12 |
| 1 | Structure d'accueil | 13 |
| 1 | La Maison de la Télédétection | 13 |
| 2 | UMR TETIS | 14 |
| 3 | IRSTEA | 14 |
| 2 | Télédétection | 15 |
| 1 | Définition de la télédétection | 15 |
| 2 | Télédétection active et passive | 16 |
| 2.1 | Télédétection active ou capteur actif | 16 |
| 2.2 | Télédétection passive ou capteur passif | 16 |
| 3 | Etat de l'art sur le LiDAR Bathymétrique et altimétrique | 17 |
| 1 | LiDAR | 18 |
| 1.1 | Définition d'un LiDAR | 18 |
| 1.2 | Principe de fonctionnement d'un LiDAR | 18 |
| 2 | LiDAR bathymétrique | 20 |
| 2.1 | Principe de fonctionnement d'un LiDAR bathymétrique | 20 |
| 2.2 | Forme d'onde bathymétrique | 21 |
| 3 | LiDAR altimétrique | 22 |
| 3.1 | Principe de fonctionnement d'un LiDAR altimétrique | 22 |
| 3.2 | Forme d'onde altimétrique | 22 |
| 4 | Wa-LID : Simulateur de forme d'onde LiDAR bathymétrique | 23 |
| 1 | Trajet d'une impulsion laser émis par un système LiDAR bathymétrique | 24 |
| 2 | Puissance totale rétrodiffusée $P_{tot}(t)$ détectée par le télescope | 25 |
| 3 | Impulsion laser émise par le système LiDAR bathymétrique | 26 |
| 4 | Puissance rétrodiffusée par la surface de l'eau $P_s(t)$ détectée par le télescope | 31 |
| 5 | Puissance rétrodiffusée $P_c(t)$ par la colonne d'eau détectée par le télescope | 37 |
| 6 | Puissance rétrodiffusée par le fond de l'eau $P_b(t)$ détectée par le télescope | 40 |
| 7 | Sources de bruit détectées par le télescope | 40 |
| 7.1 | Bruit dû à la radiation solaire détecté par le télescope $P_{bg}(t)$ | 40 |
| 7.2 | Bruit du détecteur détecté par le télescope $P_N(t)$ | 41 |
| 8 | Exemple d'une forme d'onde simulée avec le simulateur Wa-LID 2.0 | 42 |
| 9 | Intérêts du simulateur bathymétrique Wa-LID 2.0 | 43 |

| | | |
|------------|---|-----------|
| II | Les données LiDAR utilisées | 44 |
| 5 | Description des données | 45 |
| 1 | Formes d'onde bathymétriques | 46 |
| 1.1 | Formes d'onde réelles : Site du Golfe du Morbihan | 46 |
| 1.2 | Formes d'onde simulées pour deux configurations du milieu | 48 |
| 2 | Formes d'onde altimétriques | 49 |
| 2.1 | Formes d'onde réelles : Forêt du Brésil | 49 |
| 6 | Préparation des données LiDAR | 50 |
| 1 | Partie utile W_{ext} d'une forme d'onde LiDAR | 51 |
| 1.1 | Partie utile W_{ext} d'une forme d'onde LiDAR | 51 |
| 1.2 | Intérêts de calculer la partie utile d'une forme d'onde LiDAR | 51 |
| 1.3 | Comment calculer la partie utile d'une forme d'onde LiDAR ? | 51 |
| 2 | Données LiDAR après préparation | 54 |
| III | Les différentes méthodes d'inversion des formes d'onde LiDAR | 55 |
| 7 | Vers de nouvelles méthodes d'inversion de formes d'onde LiDAR | 56 |
| 1 | Les méthodes d'inversion des formes d'onde LiDAR rencontrées dans la littérature et leurs limites | 57 |
| 1.1 | Etat de l'art des méthodes d'inversion de formes d'onde LiDAR dans la littérature | 57 |
| 1.2 | Les limites des méthodes d'inversion de formes d'onde actuelles | 58 |
| 2 | Vers de nouvelles méthodes d'inversion de formes d'onde LiDAR | 59 |
| 8 | Inversion des formes d'onde LiDAR avec l'Analyse en Composantes Principales | 61 |
| 1 | Rappel sur l'Analyse en Composantes Principales | 62 |
| 1.1 | Principe de l'Analyse en Composantes Principales | 62 |
| 1.2 | Type de données pour effectuer une Analyse en Composantes Principales | 62 |
| 1.3 | Les Composantes Principales | 63 |
| 1.4 | Résumé sur l'Analyse en Composantes Principales | 64 |
| 2 | Inversion des formes d'onde LiDAR à l'aide de l'Analyse en Composantes Principales | 65 |
| 9 | Inversion des formes d'onde LiDAR avec l'analyse en ondelettes | 73 |
| 1 | Rappel : Ondelettes de Haar | 74 |
| 1.1 | Principe de l'analyse en ondelettes | 74 |
| 1.2 | La base de Haar | 74 |
| 1.3 | Les coefficients d'ondelettes | 76 |
| 1.4 | Reconstruction d'un signal de forme d'onde à l'aide des ondelettes de Haar | 76 |
| 2 | Inversion des formes d'onde LiDAR avec l'analyse en ondelettes | 78 |
| 10 | Inversion des formes d'onde LiDAR avec les splines cubiques | 81 |
| 1 | Rappel sur les splines cubiques | 82 |
| 1.1 | Principe des splines cubiques | 82 |
| 1.2 | Les coefficients des splines cubiques | 83 |
| 1.3 | Interpolation d'une forme d'onde à l'aide des polynômes cubiques par morceaux | 84 |
| 2 | Inversion des formes d'onde à l'aide des splines cubiques | 85 |

| | |
|--|------------|
| 11 Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques | 87 |
| 1 Rappel sur les B-Splines cubiques | 88 |
| 1.1 Principe des B-Splines cubiques | 88 |
| 1.2 Les fonctions de base | 89 |
| 1.3 Les coefficients des B-Splines cubiques | 90 |
| 2 Reconstruction d'une forme d'onde réelle avec les B-Splines cubiques | 91 |
| 3 Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques | 92 |
| | |
| IV Résultats obtenus et Conclusion | 94 |
| | |
| 12 Résultats et Analyses | 95 |
| 1 Résultats pour les données LiDAR bathymétriques du Golfe du Morbihan | 95 |
| 1.1 Résultats pour la base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode (APD) | 95 |
| 1.2 Résultats pour la base de données du Golfe du Morbihan avec le capteur PhotoMultiplificateur (PMT) | 97 |
| 2 Résultats pour les données LiDAR acquises avec le simulateur Wa-LID | 100 |
| 2.1 Résultats pour les données LiDAR avec une eau claire | 100 |
| 2.2 Résultats pour les données LiDAR avec une eau turbide | 103 |
| 3 Résultats pour les données LiDAR altimétriques du Brésil | 106 |
| | |
| 13 Conclusion | 109 |
| | |
| V ANNEXES | 110 |
| | |
| A Le logiciel R | 111 |
| B Simulateur Wa-LID 2.0 | 112 |
| C Reconstruire un signal à l'aide des ondelettes de Haar | 118 |
| D Inversion des formes d'onde LiDAR à l'aide de l'Analyse en Composantes Principales | 121 |
| E Inversion des formes d'onde LiDAR à l'aide de l'analyse en ondelettes | 123 |
| F Inversion des formes d'onde LiDAR à l'aide des splines cubiques | 125 |
| G Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques | 127 |
| H Validation croisée 5-folds | 129 |
| I Stepwise : Méthode de sélection automatique de variables explicatives | 130 |
| J CART | 131 |
| K Root Mean Square Error (RMSE) | 132 |
| L Les composantes principales | 133 |
| M Fonction d'échelle père ϕ et l'ondelette mère ψ | 135 |
| N Les coefficients des splines cubiques | 136 |

Table des figures

| | | |
|------|---|----|
| 2.1 | Principe de la télédétection (Référence :[65]) | 15 |
| 2.2 | Télédétection active (Référence [69]) | 16 |
| 2.3 | Télédétection passive (Référence [69]) | 16 |
| 3.1 | Système LiDAR (Référence [73]) | 18 |
| 3.2 | Signal d'un LiDAR altimétrique et bathymétrique (Référence [74]). | 19 |
| 3.3 | Coefficient d'atténuation de l'eau en fonction de la longueur d'onde (Référence [75]) | 20 |
| 3.4 | Principe du LiDAR bathymétrique (Référence [33]) | 20 |
| 3.5 | Allure générale d'une forme d'onde bathymétrique (Référence [76]) | 21 |
| 3.6 | Principe du LiDAR altimétrique (Référence [78]) | 22 |
| 3.7 | Forme d'onde altimétrique (Référence [77]) | 22 |
| 4.1 | Cheminement des trains d'onde (Référence [33]) | 24 |
| 4.2 | Propagation d'une impulsion laser par un système LiDAR bathymétrique (Référence [17]) | 25 |
| 4.3 | Décomposition d'une forme d'onde en plusieurs termes de puissance (sans les composantes du bruit. Référence : Lessaignoux [33]) | 26 |
| 4.4 | Stretching du pulse à la surface de l'eau. La courbe bleu représente la gaussienne dans l'air. La courbe rose représente l'étalement de la gaussienne à la surface de l'eau. τ_s est la différence entre la largeur à mi-hauteur de la gaussienne dans l'air et la largeur à mi-hauteur à la surface de l'eau $T_1 - T_0 = \tau_s$. | 28 |
| 4.5 | Étirement du pulse ("Pulse stretching") | 29 |
| 4.6 | Avec une pente à la surface de l'eau. | 32 |
| 4.7 | Sans pente à la surface de l'eau. | 32 |
| 4.8 | θ_t Angle de réfraction air/eau | 33 |
| 4.9 | L'émetteur et le récepteur du LiDAR sont dans la même direction. La lumière émise ou réfléchie peut-être bloquée ou non par une autre microfacette | 34 |
| 4.10 | Modèle statistique de la surface de l'eau (Référence [80]) | 34 |
| 4.11 | Distribution de Beckmann avec (a) $r=0.2$ et (b) $r=0.6$ (Référence [80]) | 35 |
| 4.12 | Les profondeurs des couches de la colonne d'eau. | 37 |
| 4.13 | Produit de convolution. Impulsion gaussienne émise par le LiDAR convoluée avec la réponse impulsionnelle de chaque couche de la colonne d'eau | 39 |
| 4.14 | Exemple d'une forme d'onde simulée avec le simulateur Wa-LID 2.0 | 42 |
| 4.15 | Simulation de deux bases de données de 780 formes d'onde LiDAR | 43 |
| 5.1 | Le Golfe du Morbihan en Bretagne sud (Référence : Tristan et al). | 46 |
| 5.2 | Représentation de deux formes d'onde contenues dans la base de données du Golfe du Morbihan. La bathymétrie mesurée sur le terrain pour les forme d'onde n°1 et n°166 est respectivement de 92,88 cm et 2.4 m. | 47 |
| 5.3 | Photo d'une parcelle d'Eucalyptus en cours de coupe ©Cirad | 49 |

| | | |
|------|---|----|
| 6.1 | Partie utile d'une forme d'onde LiDAR. | 51 |
| 6.2 | Etape 1 : la sélection du bruit | 52 |
| 6.3 | Etape 2 : Calcul de la puissance moyenne du bruit : ligne en rouge | 52 |
| 6.4 | Etape 3 : Seuil délimitant le bruit parasite du signal : ligne verte | 53 |
| 6.5 | Etape 4 : Calcul de la partie utile d'une forme d'onde LiDAR | 53 |
| 7.1 | Méthode d'inversion classique par détection de pics | 57 |
| 7.2 | On ne distingue plus la composante de surface (courbe rouge) et celle du fond de l'eau (courbe verte) (Référence [76]) | 58 |
| 7.3 | Méthodologie pour les nouvelles méthodes d'inversion de formes d'onde LiDAR. Estimation de la bathymétrie et de la hauteur des arbres en utilisant une approche statistique. | 60 |
| 8.1 | Représentation des données | 62 |
| 8.2 | Graphique des individus | 64 |
| 8.3 | Tableau des données avec les p variables initiales | 64 |
| 8.4 | Tableau des données avec les p nouvelles variables. Changement de base. | 64 |
| 8.5 | Quantité d'information recueillie par les composantes principales (ou axes factoriels) | 66 |
| 8.6 | Tableau des données avec p=64 variables initiales et 189 formes d'onde. | 67 |
| 8.7 | Tableau des données après changement de base. Nous avons fait le choix de conserver seulement les 30 premières nouvelles variables. Les 5 premières composantes principales résument 99.20 % de l'information initiale du tableau. | 67 |
| 8.8 | Graphique des individus Dim 1 et 2 représentent respectivement la première et deuxième composantes principales. | 68 |
| 8.9 | Graphique des individus (ou des formes d'onde) en fonction de deux gammes de profondeur. | 69 |
| 8.10 | Régression linéaire de la forme : $Z_{est} = \alpha PC1 + \epsilon$ avec $\alpha = 0.005391$ et $\epsilon = 1.376486$ | 70 |
| 9.1 | Ondelettes de Haar - Fonction $f(t)$ décomposée en un ensemble de fonctions élémentaires discontinues avec des niveaux de résolution de plus en plus fins. | 74 |
| 9.2 | Fonction d'échelle père ϕ et ondelette mère ψ | 74 |
| 9.3 | Les fonctions de base avec des niveaux de résolution de plus en plus fin. Les fonctions de base des deux premiers niveaux ont toujours chacune 2 fonctions de base. Le premier niveau correspond toujours aux fonctions d'échelles père , elles approximent le signal grossièrement. Les autres niveaux sont des fonctions d'ondelette mère qui sont 2 fois plus nombreuses et plus petites lorsque le niveau augmente. On répète le processus jusqu'à avoir 64 fonctions de base au total. | 75 |
| 9.4 | Coefficients d'ondelettes correspondant dans l'ordre aux fonctions de base de la figure 9.3 donnée par la fonction <code>dwt()</code> du package <code>waveslim</code> pour une forme d'onde LiDAR. | 76 |
| 9.5 | Niveau 1 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients) | 77 |
| 9.6 | Niveau 1+2 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients) et de 2 fonctions d'ondelettes mères (2 coefficients) | 77 |
| 9.7 | Niveau Niveau 1+2+3 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2 et de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients) | 77 |
| 9.8 | Niveau 1+2+3+4 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2, de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients) | 77 |
| 9.9 | Niveau 1+2+3+4+5 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2, de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients) et de 16 fonctions d'ondelettes mère du niveau 5 | 77 |

| | | |
|------|--|-----|
| 9.10 | Tous les niveaux : 1+2+3+4+5+6 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2 , de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients), de 16 fonctions d'ondelettes mère du niveau 5 et de 32 fonctions d'ondelettes mère du niveau 6. | 77 |
| 9.11 | Comparaison entre la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction <code>stepwise()</code> qui a sélectionnée 14 coefficients d'ondelettes sur les 64 coefficients d'ondelettes (tous les niveaux). | 78 |
| 9.12 | Comparaison entre la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction <code>stepwise()</code> qui a sélectionnée 11 coefficients d'ondelettes sur les 32 coefficients d'ondelettes des niveaux 1,2,3,4,5 (On ne prend pas en compte le niveau 6) | 79 |
| 10.1 | Interpolation à l'aide des splines cubiques | 82 |
| 10.2 | Coefficients des polynômes cubiques par morceaux pour l'interpolation d'une forme d'onde LiDAR avec la fonction <code>interpSpline()</code> du package <code>splines</code> | 83 |
| 10.3 | Interpolation d'une forme d'onde réelle en utilisant 5, 10 ,20 et 30 polynômes cubiques | 84 |
| 10.4 | Comparaison de la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction <code>stepwise</code> qui a sélectionné 17 variables explicatives sur 120 variables explicatives (les coefficients des splines cubiques). | 85 |
| 11.1 | Coefficients des B-Splines associés à la figure 1.1 pour le cas des 5 fonctions de base | 90 |
| 11.2 | Approximation d'une forme d'onde réelle du Golfe du Morbihan avec les B-Splines cubiques. | 91 |
| 11.3 | Comparaison de la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction <code>stepwise</code> qui a sélectionné 3 variables explicatives sur 5 variables explicatives (les coefficients des B-Splines cubiques). | 92 |
| 12.1 | Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR simulées avec une eau turbide. | 104 |
| H.1 | Schéma de la validation croisée 5-folds | 129 |
| L.1 | Nuage de points. On cherche la direction de "plus grand allongement" du nuage - Référence : [80] | 133 |
| L.2 | La première composante principale (PC : " Principal component ") - Référence : [80] | 133 |
| L.3 | On cherche la direction avec le maximum de l'inertie résiduelle de sorte que PC_2 soit perpendiculaire (non corréllées) à PC_1 - Référence : [80] | 134 |
| L.4 | La deuxième composante principale (PC : " Principal component ") - - Référence : [80] | 134 |

Liste des tableaux

| | | |
|------|---|----|
| 5.1 | Tableau de 189 formes d'onde enregistrées par un LiDAR bathymétrique avec le capteur APD. | 46 |
| 5.2 | Paramètres instrumentaux | 48 |
| 5.3 | Paramètres pour une eau claire | 48 |
| 5.4 | Paramètres pour une eau plus turbide | 49 |
| 6.1 | Caractéristiques des bases de données | 54 |
| 8.1 | Tableau de 189 formes d'onde enregistrées par un lidar bathymétrique avec le capteur APD. Les lignes correspondent aux individus (Formes d'onde) et les colonnes aux variables (Puissance rétrodiffusée détectée par le LiDAR à un instant donné t_j ($j=1,\dots,64$). La colonne "bathymétrie mesurée" correspond à une variable supplémentaire . Elle n'interviendra pas dans la construction des composantes principales. Elle nous permettra de voir la corrélation avec les composantes principales et de construire des modèles statistiques permettant d'estimer la bathymétrie. | 65 |
| 8.2 | Tableau des valeurs propres, du pourcentage de variance et de la variance cumulée des axes principaux ("PC : Principal Component"). | 67 |
| 8.3 | Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables descriptifs des différents modèles mathématiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : Random Forest, stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 30 premières variables, W_{ext} est la variable qui correspond à la partie utile d'une forme d'onde LiDAR.) | 72 |
| 9.1 | Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables explicatives des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : RandomForest, stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 64,32,16,8 et 4 variables). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (Partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses. | 80 |
| 10.1 | Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables explicatives des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : RandomForest ; Stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 120, 80, 40 et 20 variables explicatives). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (Partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses. | 86 |
| 11.1 | Base de 5,10, 20 et 30 B-Splines cubiques uniformes. Une fonction de base est un polynôme de degré 3. | 89 |

| | | |
|------|--|-----|
| 11.2 | Coefficient de détermination R^2 et le RMSE des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : Random Forest ; Stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 5,10,20 et 30 variables explicatives (5, 10, 20 et 30 fonctions de base). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses. | 93 |
| 12.1 | Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques). Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode. | 96 |
| 12.2 | Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Base de données du Golfe du Morbihan avec le capteur PhotoMultiplicateur Tube. | 98 |
| 12.3 | Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR simulées avec une eau claire. | 101 |
| 12.4 | Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR altimétriques (forêt brésil). | 107 |
| 12.5 | Forêt Brésil - Les meilleurs résultats (RMSE (m)) de la figure 12.1 avec la régression linéaire multiple, CART et Random Forest avec un nombre de variables explicatives inférieur à 5 en utilisant l'Analyse en Composantes Principales, l'analyse en ondelette, les splines cubiques et les B-Splines cubiques. | 108 |
| I.1 | Tableau que l'on nomme "Table" | 130 |

Remerciements

Je remercie très chaleureusement **Nicolas BAGHDADI** (Directeur Scientifique du pôle THEIA) et **Jean-Stéphane BAILLY** (ingénieur des ponts, des eaux et forêts, AgroParisTech) pour m'avoir proposé ce sujet de stage extrêmement intéressant et enrichissant, et dont la grande compétence et sympathie ont permis le bon déroulement de celui-ci et motivé mon implication dans ce projet.

Je remercie également **Henri LABORDE** (IE CNRS UMR 5554 ISEM) pour la relecture de mon rapport.

Et enfin, je remercie aussi les stagiaires de la Maison de la Télédétection pour m'avoir apportée leur bonne humeur pendant ces six mois.

Introduction

L'utilisation des capteurs LiDAR dans les domaines environnemental connaît aujourd'hui une forte croissance comme en témoigne le nombre d'articles scientifiques sur l'application du LiDAR (plus de 1100 articles scientifiques entre 2010 et 2014).

Dans ce contexte, **l'objectif de mon stage** est d'estimer à partir de formes d'onde LiDAR réelles et simulées la bathymétrie¹ et la hauteur des arbres en utilisant de nouvelles méthodes d'inversion des formes d'onde s'appuyant sur un apprentissage statistique et sur une réduction de la dimension du signal. Pour ce faire, J'utilise dans ce stage des données réelles mesurées par des capteurs aéroportés ainsi que des données simulées numériquement afin de travailler sur des jeux de données complètement contrôlés.

Avant de pouvoir estimer la bathymétrie sur les formes d'onde simulées, j'ai été chargé de comprendre, d'améliorer, de documenter et de corriger le code source d'un simulateur bathymétrique² : Wa-LID 2.0 (Water LiDAR). Le simulateur m'a permis de générer deux bases de données de 780 formes d'onde LiDAR avec deux configurations différentes du milieu (une eau claire et une eau turbide). J'avais également à ma disposition plusieurs bases de données LiDAR réelles : les données LiDAR bathymétriques acquises en Juin 2005 sur le Golfe du Morbihan³ par le SHOM (Service Hydrographique et Océanographique de la Marine) et les données LiDAR altimétriques acquises au Brésil par le capteur GLAS (Géoscience Laser Altimeter System) embarqué sur le satellite IceSAT.

Ce rapport de stage s'articule autour de 12 chapitres. **Les chapitres (1) et (2)** présentent la structure d'accueil et le domaine de recherche dans lequel j'ai effectué mon stage. **Le chapitre (3)** fait l'état de l'art sur les LiDAR bathymétriques et altimétriques. **Le chapitre (4)** fournit les différentes équations qui ont permis de modéliser une forme d'onde avec le simulateur Wa-LID et explique l'intérêt de celui-ci. **Le chapitre (5)** décrit les données LiDAR bathymétriques et altimétriques dont je dispose. **Le chapitre (6)** est une préparation préliminaire des données LiDAR. **Le chapitre (7) est un chapitre clé** expliquant l'approche statistique que l'on va utiliser pour créer différents modèles statistiques (Régression linéaire multiple, arbre de décision et arbres aléatoires) à l'aide de **variables d'intérêts** (bathymétrie ou hauteur d'arbre suivant la base de données) et de **variables explicatives**. Dans ce rapport de stage, je mets plus l'accent sur les formes d'onde LiDAR bathymétriques mais la démarche scientifique est la même pour les formes d'onde LiDAR altimétriques. On choisit une base de données de N formes d'onde. Chaque forme d'onde est caractérisée par une bathymétrie (ou une hauteur d'arbre) mesurée sur le terrain. On détermine les variables explicatives à l'aide de quatre méthodes différentes : l'Analyse en Composantes Principales, l'analyse en ondelettes, les splines cubiques et les B-Splines cubiques. On sélectionne aléatoirement 80% de formes d'onde de la base donnée pour construire un modèle statistique (Régression linéaire multiple, arbre de décision (CART) et les arbres aléatoires (Random Forest)). Pour valider ce modèle prédictif, on l'applique sur les 20 % des formes d'onde restantes. Je répète 5 fois ce processus en utilisant la validation croisée⁴ (k-folds, $k=5$). On peut ainsi comparer la bathymétrie mesurée sur le terrain et la bathymétrie que nous avons estimée avec le modèle construit. **Les chapitres (8), (9), (10) et (11)** expliquent comment calculer les variables explicatives des

1. Définition : La bathymétrie est la science de la mesure des profondeurs et du relief de l'océan pour déterminer la topographie du sol de la mer (Wikipédia)

2. Malheureusement, on ne dispose pas encore de simulateur altimétrique

3. à l'aide de deux capteurs LiDAR différents

4. cf. Annexe H

formes d'onde LiDAR en utilisant l'Analyse en Composantes Principales, l'analyse en ondelettes, les splines cubiques et les B-Splines cubiques. Les méthodes utilisées sont seulement des "outils" permettant d'extraire les variables explicatives. Et enfin, **le chapitre (12)** présente les résultats et analyses des différents modèles statistiques. On utilise deux critères qui définissent la précision des modèles statistiques : le coefficient de détermination R^2 (en %) et le RMSE⁵ (en m) (**R**oot **M**ean **S**quarre **E**rreur).

5. Erreur quadratique moyenne

Première partie

La technologie LiDAR : Etat de l'art

Chapitre 1

Structure d'accueil

Sommaire

| | | |
|---|--|----|
| 1 | La Maison de la Télédétection | 13 |
| 2 | UMR TETIS | 14 |
| 3 | IRSTEA | 14 |

Résumé du chapitre : Ce chapitre fournit une présentation générale de la structure qui m'a accueillie durant mon stage.

J'ai effectué mon stage à la **Maison de la Télédétection** en Languedoc-Roussillon au sein de **l'unité mixte de recherche TETIS** dans l'équipe **ATTOS**¹ sous contrat avec **l'IRSTEA** en sous-encadrement mixte avec IRSTEA et AgroParisTech. Il a été réalisé du 4 Mars 2014 au 31 Août 2014 sous la tutelle de Nicolas BAGHDADI et Jean Stéphane BAILLY.

1 La Maison de la Télédétection

La Maison de la Télédétection en Languedoc-Roussillon conduit des recherches appliquées et des formations de haut niveau en Télédétection et Information Géographique pour la gestion de l'agriculture, de l'environnement et des territoires. Localisée au sein du campus d'Agropolis à Montpellier, elle regroupe une centaine de personnes issues d'équipes regroupées dans deux UMR : UMR TETIS et UMR Espace-Dev.



1. Acquisition, Télédétection, Traitement et Observations Spatiales

2 UMR TETIS

Ce stage de fin d'études du Master Physique Informatique de l'Université Montpellier 2 s'est déroulé à la Maison de la Télédétection au sein de l'Unité Mixte de Recherche Territoires, Environnement, Télédétection et Information Spatiale (UMR TETIS) qui est une UMR entre AgroParisTech, Cirad et IRSTEA.

TETIS regroupe quatre équipes de recherche :

- **ATTOS** : Acquisition, Télédétection, Traitement et Observations Spatiales.
- **AMoS** : Analyse et Modélisation Spatiales.
- **SISO** : Système d'Information Spatialisée : modélisation, extraction et diffusion des données et connaissances.
- **UsIG** : Usage de l'Information spatiale et Gouvernance.



3 IRSTEA

Mon stage a été financé par l'Institut national de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA) dans le cadre d'un projet CNES-TOSCA, le projet "BD SOBLA" entre l'UMR TETIS, L'Institut universitaire Européen de la Mer (Université de Bretagne occidentale), L'IFREMER et le SHOM. IRSTEA est un établissement public à caractère scientifique et technologique sous la double tutelle des ministères en charges de la Recherche et de l'Agriculture.

IRSTEA est un établissement public à caractère scientifique et technologique (EPST). Réparties en France dans 9 centres, ses activités de recherches et d'expertises sont tournées vers l'action et l'appui aux politiques publiques.



Chapitre 2

Téledétection

Sommaire

| | | |
|----------|---|-----------|
| 1 | Définition de la télédétection | 15 |
| 2 | Téledétection active et passive | 16 |
| 2.1 | Téledétection active ou capteur actif | 16 |
| 2.2 | Téledétection passive ou capteur passif | 16 |

Résumé du chapitre : Dans ce chapitre, je présente rapidement le domaine de la télédétection qui peut-être divisé en deux sous-catégories : La télédétection active et passive.

1 Définition de la télédétection

La télédétection [66] désigne **la mesure** ou l'acquisition d'informations **sur un objet** ou un phénomène, **par l'intermédiaire d'un instrument de mesure n'ayant pas de contact avec l'objet étudié**. C'est l'utilisation à distance de n'importe quel type d'instrument permettant l'acquisition d'informations sur l'environnement.

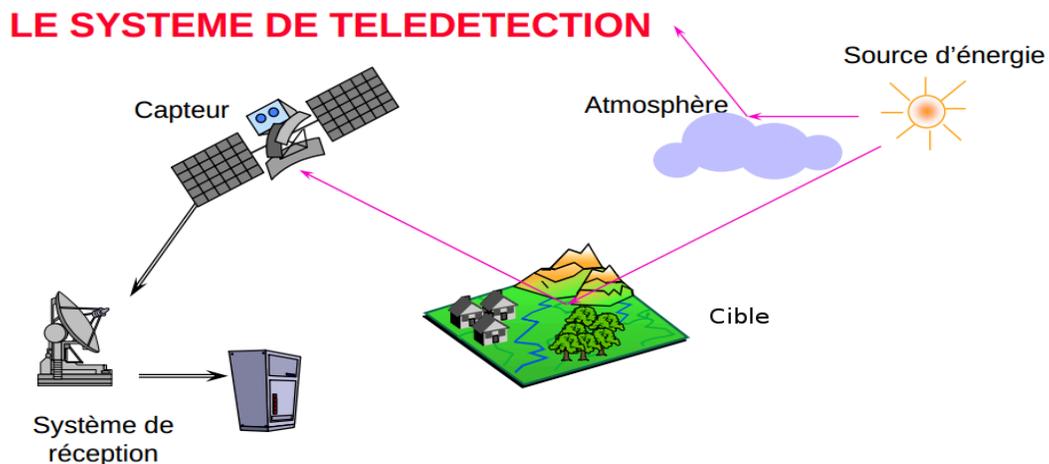


FIGURE 2.1 – Principe de la télédétection (Référence :[65])

La télédétection moderne se rapporte à la détection d'objets par l'intermédiaire d'ondes électromagnétiques. La télédétection [67] regroupe 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 (Figure 2.1). L'outil qui nous permet d'acquérir le rayonnement par une cible est donc un capteur, et celui-ci est placé sur un dispositif qui lui permet d'observer la cible à distance (avion, satellite...).

2 Télédétection active et passive

On distingue **télédétection active et passive** [68], selon que la source de **rayonnement est artificielle, ou naturelle**.

2.1 Télédétection active ou capteur actif

Le **capteur actif** [68] embarqué à bord de l'avion ou du satellite **produit sa propre énergie** pour illuminer la cible : il dégage un rayonnement électromagnétique qui est dirigé vers la cible. Le rayonnement réfléchi par la cible est alors perçu et mesuré par le capteur. On utilise des radars ou des lasers comme source d'énergie (Figure 2.2).

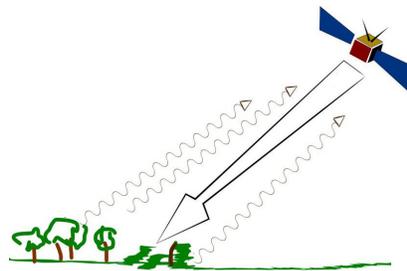


FIGURE 2.2 – Télédétection active (Référence [69])

2.2 Télédétection passive ou capteur passif

Le **capteur passif** [68] embarqué sur l'avion ou le satellite **capte le rayonnement réfléchi par la surface de la Terre**, lorsque le Soleil illumine la Terre. **La source d'énergie est naturelle et n'est donc pas produite par le capteur**. Le Soleil est la source d'énergie. Il n'est donc pas possible de réaliser des observations la nuit ou quand des nuages masquent le sol (Figure 2.3 et 2.1).

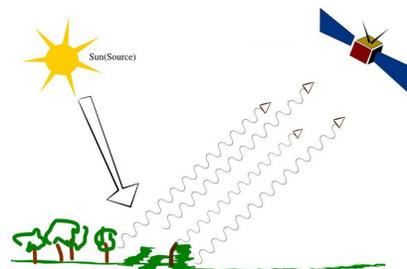


FIGURE 2.3 – Télédétection passive (Référence [69])

Chapitre 3

Etat de l'art sur le LiDAR Bathymétrique et altimétrique

Sommaire

| | | |
|----------|---|-----------|
| 1 | LiDAR | 18 |
| 1.1 | Définition d'un LiDAR | 18 |
| 1.2 | Principe de fonctionnement d'un LiDAR | 18 |
| 2 | LiDAR bathymétrique | 20 |
| 2.1 | Principe de fonctionnement d'un LiDAR bathymétrique | 20 |
| 2.2 | Forme d'onde bathymétrique | 21 |
| 3 | LiDAR altimétrique | 22 |
| 3.1 | Principe de fonctionnement d'un LiDAR altimétrique | 22 |
| 3.2 | Forme d'onde altimétrique | 22 |

Résumé du chapitre : Dans ce chapitre, je fais l'état de l'art sur le LiDAR bathymétrique et altimétrique.

1 LiDAR

1.1 Définition d'un LiDAR

L'acronyme **LiDAR** signifie **Ligh Detection And Ranging**. C'est un **système actif** (Figure 2.2) qui fonctionnent sur le même principe qu'un radar ordinaire, à l'exception du fait que des impulsions d'ondes optiques à haute résolution sont émises par une source laser, au lieu des ondes radio. Il opère dans le domaine spectral du **visible**, **UltraViolet** et **InfraRouge** (10^{-4} à 10^{-8} m).

Un **LiDAR** [71] (Figure 3.1) se compose :

- d'un **système laser** chargé d'émettre l'onde lumineuse (émetteur)
- d'un **télescope** qui collecte l'onde rétrodiffusé (récepteur)
- d'une **chaîne de traitement** qui quantifie le signal reçu.

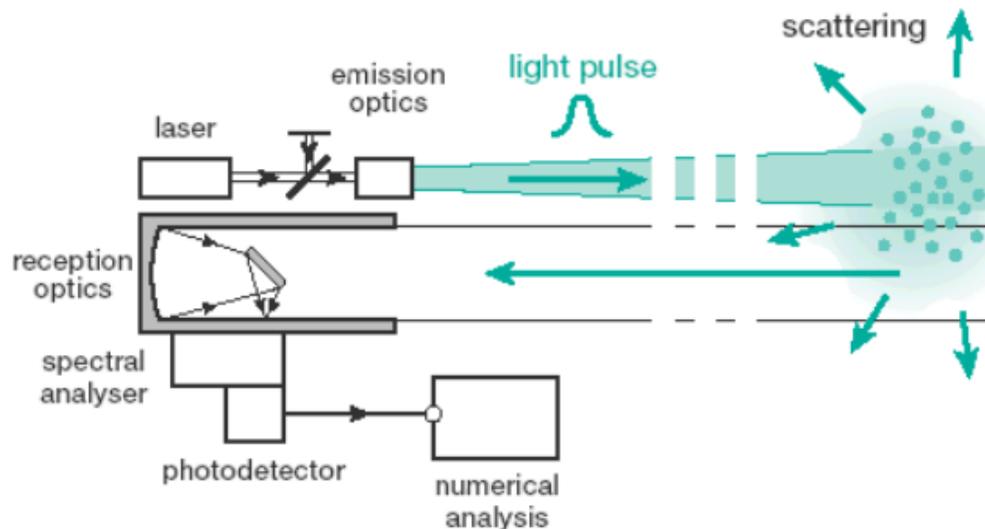


FIGURE 3.1 – Système LiDAR (Référence [73])

1.2 Principe de fonctionnement d'un LiDAR

Le système laser [72] génère (Figure 3.1), à **intervalles réguliers**, une **série d'impulsions lumineuses** intenses, à **une fréquence choisie** en fonction de **l'information recherchée**. Ces impulsions sont émises en direction de l'objet d'étude (surface solide, liquide ou gazeuse). En interagissant avec la surface, le signal se scinde en plusieurs composantes :

- l'une est **diffusée ou absorbée** par le milieu
- l'autre est **réfléchi**e et en partie renvoyée en direction de la source du rayonnement laser.

Le **signal rétrodiffusé** est recueilli par le détecteur, filtré puis enregistré pour être analysé car il contient un certain nombre d'informations sur le **milieu et ses caractéristiques physiques**.

De plus, l'**écart en temps** entre **l'instant d'émission** et **l'instant de réception du signal** est représentatif de la **distance parcourue par le laser**, donc de la distance séparant la source de rayonnement de l'objet.

Dans le cadre de mon travail, je me suis intéressé à deux types de LiDAR :

- le **LiDAR bathymétrique** : pour estimer la profondeur des eaux
- le **LiDAR altimétrique** : pour calculer la hauteur des arbres.

La figure 3.2 illustre le signal reçu par ces deux types de LiDAR. Le **signal rétro-diffusé** (puissance en fonction du temps) qui est recueilli par le détecteur est communément appelé "**forme d'onde**" ou "**train d'onde**".

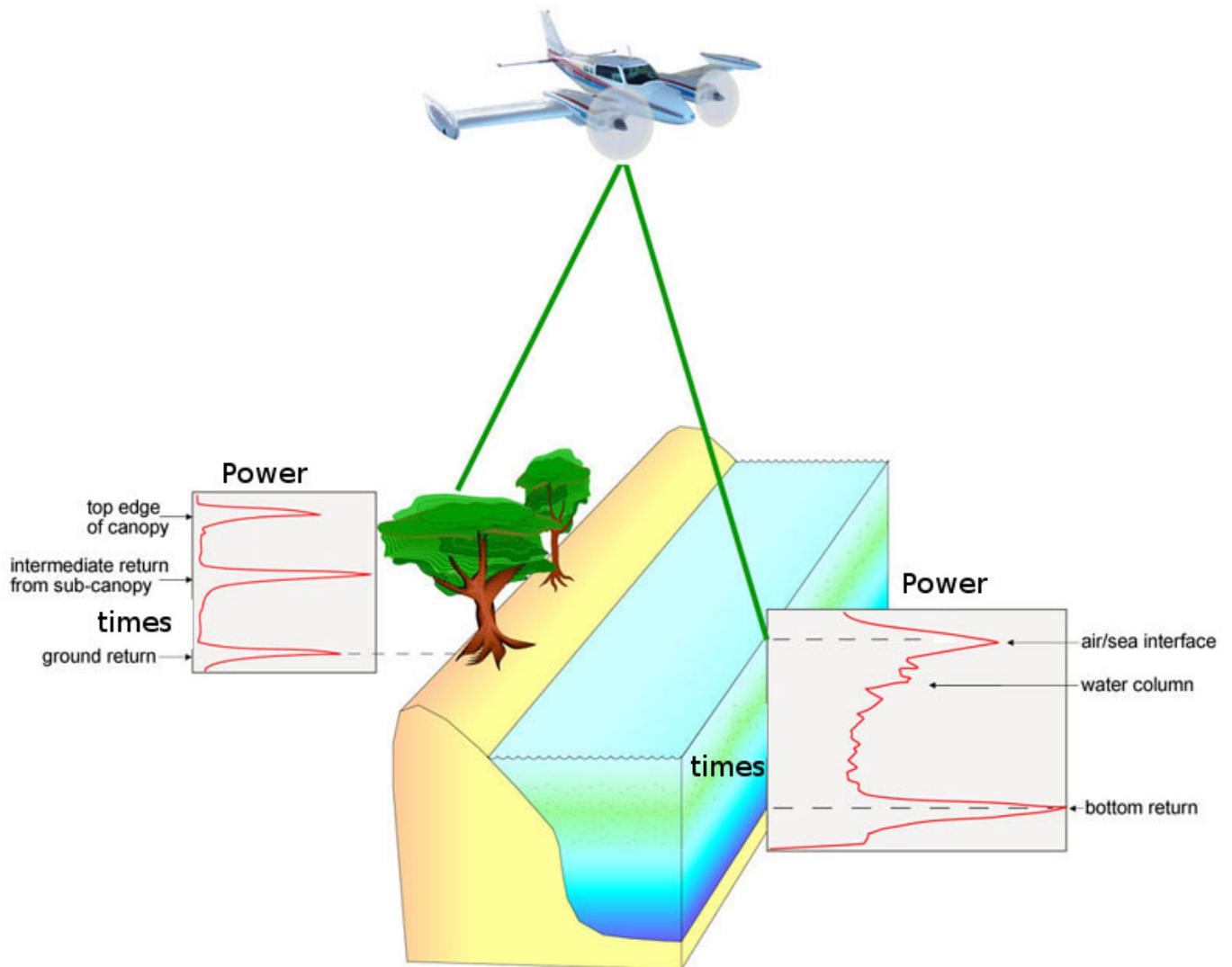


FIGURE 3.2 – Signal d'un LiDAR altimétrique et bathymétrique (Référence [74]).

2 LiDAR bathymétrique

2.1 Principe de fonctionnement d'un LiDAR bathymétrique

A la différence du LiDAR altimétrique, le **LiDAR bathymétrique** [33] émet un laser avec **une longueur d'onde** qui doit être **capable de pénétrer l'eau jusqu'au fond de l'eau**. La longueur d'onde utilisée pour le LiDAR altimétrique se situe dans l'infrarouge, or celle-ci ne pénètre dans l'eau que de quelques centimètres. Il faut donc placer sur le système un autre laser capable de traverser la colonne d'eau et d'être réfléchi par le fond de l'eau. Pour cela la longueur d'onde la plus appropriée est **le vert**. Les systèmes qui ont été développés utilisent couramment une longueur d'onde de 532 nm car elle pénètre mieux l'eau avec un minimum d'atténuation (Figure 3.3).

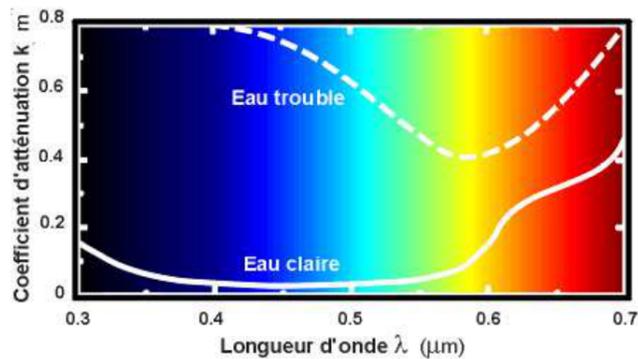


FIGURE 3.3 – Coefficient d'atténuation de l'eau en fonction de la longueur d'onde (Référence [75])

Les impulsions sont émises typiquement dans 2 longueurs d'ondes, **l'infrarouge (1064 nm) qui ne pénètre quasiment pas l'eau et est réfléchi et absorbé par la surface de l'eau** et **le vert (532 nm) qui pénètre l'eau et est réfléchi par la surface de l'eau, la colonne d'eau et le fond de l'eau**. Ainsi l'écart de temps qui sépare l'arrivée du signal infrarouge réfléchi par la surface de l'eau et celle du signal réfléchi par le fond de l'eau est fonction de la profondeur. Sachant que l'on connaît la vitesse de propagation de la lumière dans l'air comme dans l'eau, il nous est possible de déterminer la bathymétrie à un endroit donné par le système LiDAR (Figure 3.4) : $Z = c_{eau} \Delta t = c_{eau}(t_2 - t_1)$.

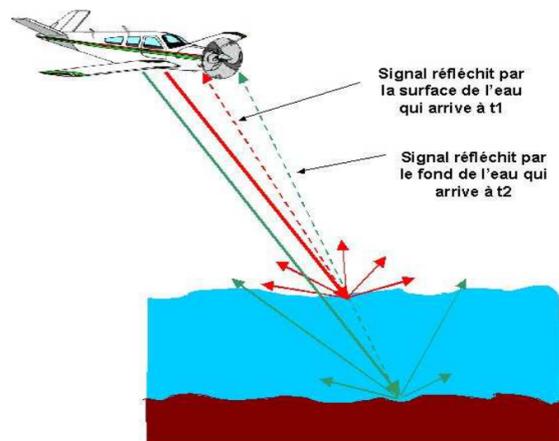


FIGURE 3.4 – Principe du LiDAR bathymétrique (Référence [33])

2.2 Forme d'onde bathymétrique

La figure 3.5 représente la puissance reçue en fonction du temps pour un système LiDAR bathymétrique. Le train d'onde bathymétrique possède généralement trois composantes "bien distinctes" :

- la première correspond au retour lié à l'interface air/eau
- la deuxième composante peut-être différenciée : c'est un signal plus étalé, moins amplifié, venant de phénomènes de rétrodiffusion au sein de la colonne d'eau
- la troisième au retour du fond de l'eau.

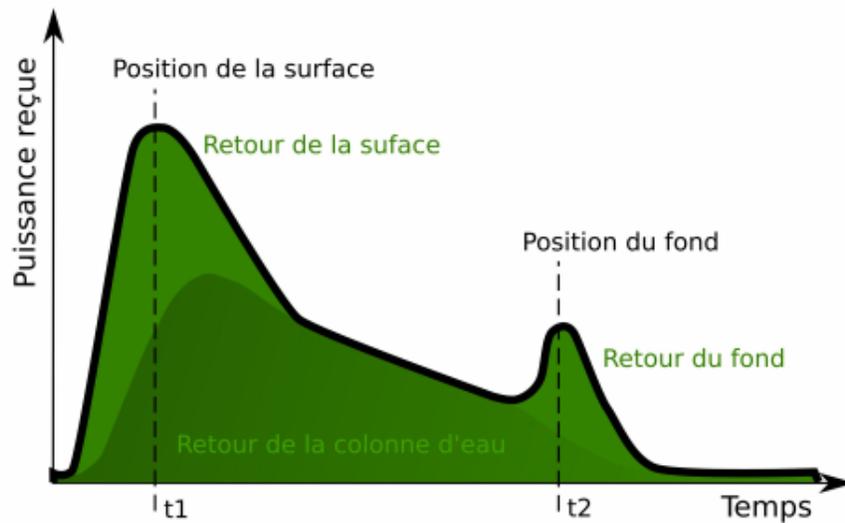


FIGURE 3.5 – Allure générale d'une forme d'onde bathymétrique (Référence [76])

3 LiDAR altimétrique

3.1 Principe de fonctionnement d'un LiDAR altimétrique

Le **LiDAR altimétrique** (Figure 3.6) émet un laser avec une longueur d'onde qui doit être capable de traverser les branches d'arbres et être réfléchi par le sol. La longueur d'onde utilisée pour le LiDAR altimétrique se situe dans l'**infrarouge**. Il fonctionne sur le même principe que le LiDAR bathymétrique. La surface de l'eau est remplacée par le haut de l'arbre, la colonne d'eau par les feuillages de l'arbre et le fond de l'eau par le sol.

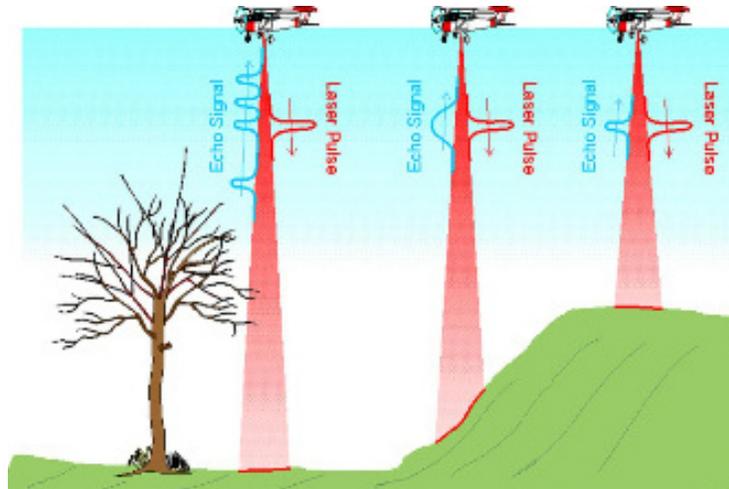


FIGURE 3.6 – Principe du LiDAR altimétrique (Référence [78])

3.2 Forme d'onde altimétrique

La figure 3.7 représente la puissance reçue en fonction du temps pour un système LiDAR altimétrique. Le train d'onde altimétrique possède généralement plusieurs composantes :

- Le premier pic correspond au retour lié à la hauteur de l'arbre
- Des pics intermédiaires entre le premier et dernier pic correspondant aux obstacles entre le haut de l'arbre et le sol
- Le dernier pic correspond au retour lié au sol.

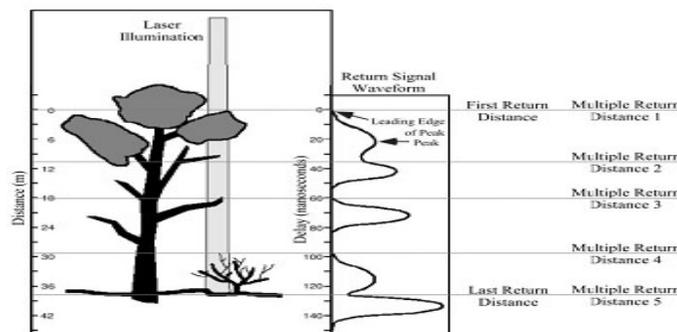


FIGURE 3.7 – Forme d'onde altimétrique (Référence [77])

Chapitre 4

Wa-LID : Simulateur de forme d'onde LiDAR bathymétrique

Sommaire

| | | |
|-----|--|----|
| 1 | Trajet d'une impulsion laser émis par un système LiDAR bathymétrique . . . | 24 |
| 2 | Puissance totale rétrodiffusée $P_{tot}(t)$ détectée par le télescope | 25 |
| 3 | Impulsion laser émise par le système LiDAR bathymétrique | 26 |
| 4 | Puissance rétrodiffusée par la surface de l'eau $P_s(t)$ détectée par le télescope | 31 |
| 5 | Puissance rétrodiffusée $P_c(t)$ par la colonne d'eau détectée par le télescope . | 37 |
| 6 | Puissance rétrodiffusée par le fond de l'eau $P_b(t)$ détectée par le télescope . | 40 |
| 7 | Sources de bruit détectées par le télescope | 40 |
| 7.1 | Bruit dû à la radiation solaire détecté par le télescope $P_{bg}(t)$ | 40 |
| 7.2 | Bruit du détecteur détecté par le télescope $P_N(t)$ | 41 |
| 8 | Exemple d'une forme d'onde simulée avec le simulateur Wa-LID 2.0 | 42 |
| 9 | Intérêts du simulateur bathymétrique Wa-LID 2.0 | 43 |

Résumé du chapitre : A travers ce chapitre, je vais faire la revue des équations permettant de générer une (ou plusieurs) forme(s) d'onde LiDAR avec le simulateur Wa-LID 2.0 (**Water-LiDAR 2.0**). Il est le résultat de plusieurs travaux d'étudiants précédents et successifs : Audrey Lesaignoux, Tristan Allouis, Hani Abdallah, Lydia Abady, Anis Bouhdaoui et Miloud Mezian.

Le simulateur intègre :

- (i) une représentation géométrique explicite des propriétés de la surface de l'eau,
- (ii) un modèle de transfert radiatif dans l'eau en fonction des propriétés physico-chimiques de l'eau et de la longueur d'onde,
- (iii) une modélisation des deux sources de bruit que sont le bruit dû au rayonnement solaire et le bruit interne du détecteur LiDAR.

Il est capable de :

- simuler un train d'onde rétro-diffusé par une cible avec une épaisseur d'eau, pour une longueur d'onde allant de 400 à 700 nm.
- simuler un train d'onde rétrodiffusé par une surface d'eau, (sans pénétration) entre 900 et 1500 nm.

1 Trajet d'une impulsion laser émis par un système LiDAR bathymétrique

Le cheminement du laser [33] subit des modifications à son passage dans la colonne d'air, à l'interface air/eau, dans la colonne d'eau et au fond de l'eau. En observant la figure 4.1 on peut décomposer le cheminement du train d'onde en 4 parties :

- A travers **la colonne d'air**, le train d'onde subit une atténuation due à l'atmosphère. Le laser émettant suivant un certain angle par rapport au Nadir et suivant un certain diamètre, forme à la surface de l'eau une empreinte, dont le diamètre est représenté par le spot "vert" sur la figure 4.1.
- A **l'interface air/eau**, le rayon laser subit une réfraction, et celui-ci va s'atténuer exponentiellement dans **la colonne d'eau** du fait des phénomènes de diffusion et d'absorption. Cette atténuation entraîne une augmentation de la taille du spot au fond de l'eau (spot "bleu" sur la figure 4.1).
- Une fois **au fond de l'eau**, le train d'onde est réfléchi et traverse à nouveau la colonne d'eau. A la surface de l'eau, celui-ci possède un diamètre (spot "rose") plus important que le spot "vert" de départ.
- La détection de ce train d'onde s'effectue avec un certain champ de vue (FOV¹) du télescope. Il n'y a alors qu'une partie de l'énergie laser retrodiffusée qui est détectée par le télescope.

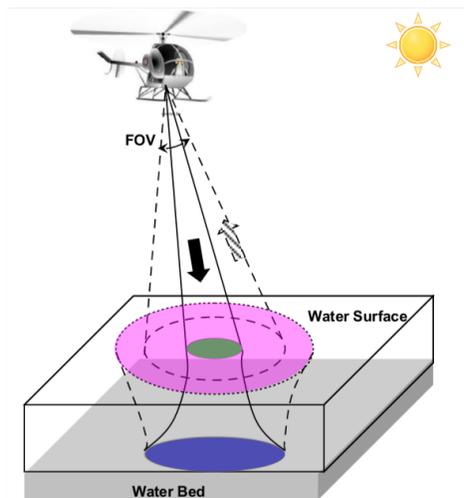


FIGURE 4.1 – Cheminement des trains d'onde (Référence [33])

1. field of view

2 Puissance totale rétrodiffusée $P_{tot}(t)$ détectée par le télescope

Le signal LiDAR reçu $P_{tot}(t)$ en fonction du temps s'écrit comme une somme de différentes contributions correspondant à :

$$P_{tot}(t) = P_s(t) + P_c(t) + P_b(t) + P_{bg}(t) + P_N(t) \quad (4.1)$$

où :

- $P_s(t)$ est la puissance reçue sur le détecteur de la surface de l'eau (section 4).
- $P_c(t)$ est la puissance reçue sur le détecteur de la colonne l'eau (section 5).
- $P_b(t)$ est la puissance reçue sur le détecteur du fond de l'eau (section 6).
- $P_{bg}(t)$ est la puissance reçue sur le détecteur de la colonne d'air (lumière naturelle, atmosphère,...) (section 7.1).
- $P_N(t)$ est la puissance reçue sur le détecteur du bruit du détecteur (section 7.2).

La puissance totale se décompose donc en 5 termes. Tous ces termes sont représentés sur la figure (4.2 et 4.3)

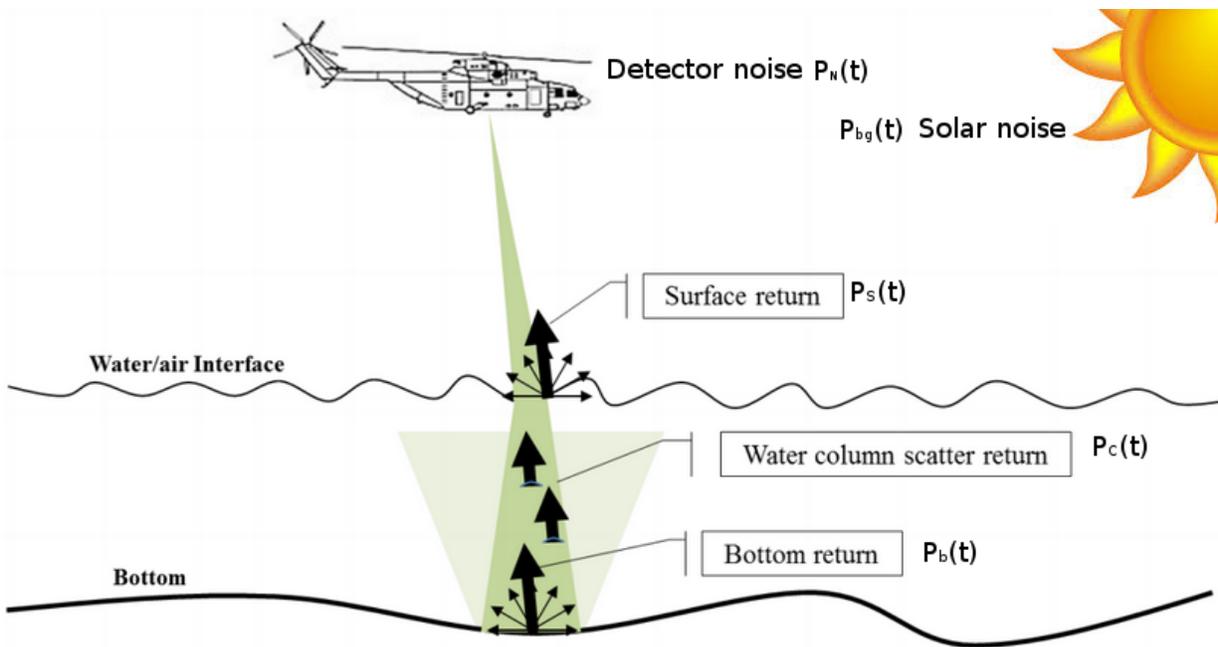


FIGURE 4.2 – Propagation d'une impulsion laser par un système LiDAR bathymétrique (Référence [17])

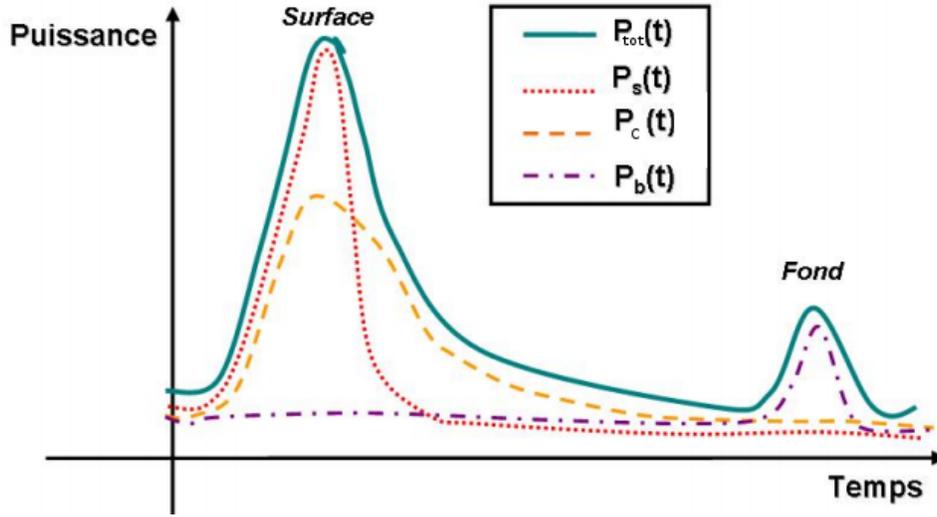


FIGURE 4.3 – Décomposition d’une forme d’onde en plusieurs termes de puissance (sans les composantes du bruit. Référence : Lessaignoux [33])

Nous allons à présent donner les équations de ces différentes puissances qui ont permis la modélisation de la forme d’onde.

3 Impulsion laser émise par le système LiDAR bathymétrique

Le pulse laser émis par le système LIDAR est considéré comme une distribution gaussienne de la forme (Lessaignoux [33]) :

$$w(t, tx, \sigma_x) = \frac{2P_e}{\sqrt{2\pi}} e^{-\frac{(t-t_x)^2}{2\sigma_x^2}} \quad (4.2)$$

où :

- t (s) est la variable temporelle.
- P_e (W) est la puissance instantanée émise par le laser.
- σ_x (s) est :
 - soit l’écart-type de la gaussienne dans l’air σ
 - soit l’écart-type de la gaussienne dans l’eau σ_1
 - ou soit l’écart-type de la gaussienne au fond de l’eau σ_2 .
- t_x (s) est :
 - soit le temps aller/retour entre le détecteur et la surface de l’eau t_s .
 - soit le temps aller/retour entre le détecteur et la colonne d’eau t_c .
 - ou soit le temps aller/retour entre le détecteur et le fond de l’eau t_b .

Puissance émise par le laser P_e

La puissance émise par le laser est donnée par :

$$P_e = \frac{E_0}{T_0} \quad (4.3)$$

- P_e (W) est la puissance émise par le laser.
- E_0 (J) est l'énergie émise par le laser.
- T_0 (s) est la largeur à mi-hauteur du pulse de la gaussienne dans l'air.

Ecart-type de la gaussienne dans l'air, l'eau et au fond de l'eau

L'écart-type du pulse dans l'air s'écrit :

$$\sigma = \frac{T_0}{2\sqrt{2\ln(2)}} \quad (4.4)$$

- T_0 (s) est la largeur à mi-hauteur du pulse de la gaussienne dans l'air .

L'écart-type du pulse dans l'eau s'écrit :

$$\sigma_1 = \frac{T_1}{2\sqrt{2\ln(2)}} \quad (4.5)$$

- T_1 (s) est la largeur à mi-hauteur du pulse de la gaussienne dans l'eau (section 3).

L'écart-type du pulse au fond de l'eau s'écrit :

$$\sigma_2 = \frac{T_2}{2\sqrt{2\ln(2)}} \quad (4.6)$$

- T_2 (s) est la largeur à mi-hauteur du pulse de la gaussienne au fond de l'eau (section 3).

Largeur à mi-hauteur du pulse de la gaussienne dans l'air, l'eau et au fond de l'eau

Au cours de son trajet le pulse laser subit deux "stretching" (étirement du pulse : augmentation de la largeur à mi-hauteur). Une première fois, à la **surface de l'eau** et une seconde fois **au fond de l'eau**. La figure (4.4) représente l'étirement du pulse à la surface de l'eau.

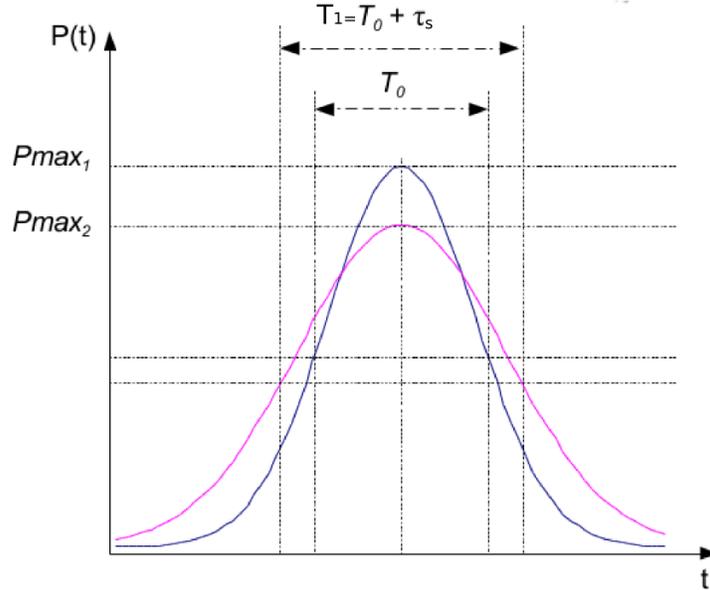


FIGURE 4.4 – Stretching du pulse à la surface de l'eau. La courbe bleu représente la gaussienne dans l'air. La courbe rose représente l'étalement de la gaussienne à la surface de l'eau. τ_s est la différence entre la largeur à mi-hauteur de la gaussienne dans l'air et la largeur à mi-hauteur à la surface de l'eau $T_1 - T_0 = \tau_s$.

En effet, lorsqu'un pulse laser est émis par un système LiDAR à une altitude H ayant **un angle d'incidence θ** (provenant soit d'une pente locale soit d'un angle de visée oblique du capteur) **et un angle d'ouverture du faisceau laser** (angle de divergence) γ , une différence de temps² entre le chemin le plus court et le plus long intervient. C'est cette différence de temps qui entraîne **l'étirement du pulse**. La figure (4.5) illustre :

- la différence de temps Δt_s entre le chemin AB et AC que parcourt le laser dans l'air.
- la différence de temps Δt_b entre le chemin BE et CG que parcourt le laser dans l'eau.

2. ou différence de chemin optique

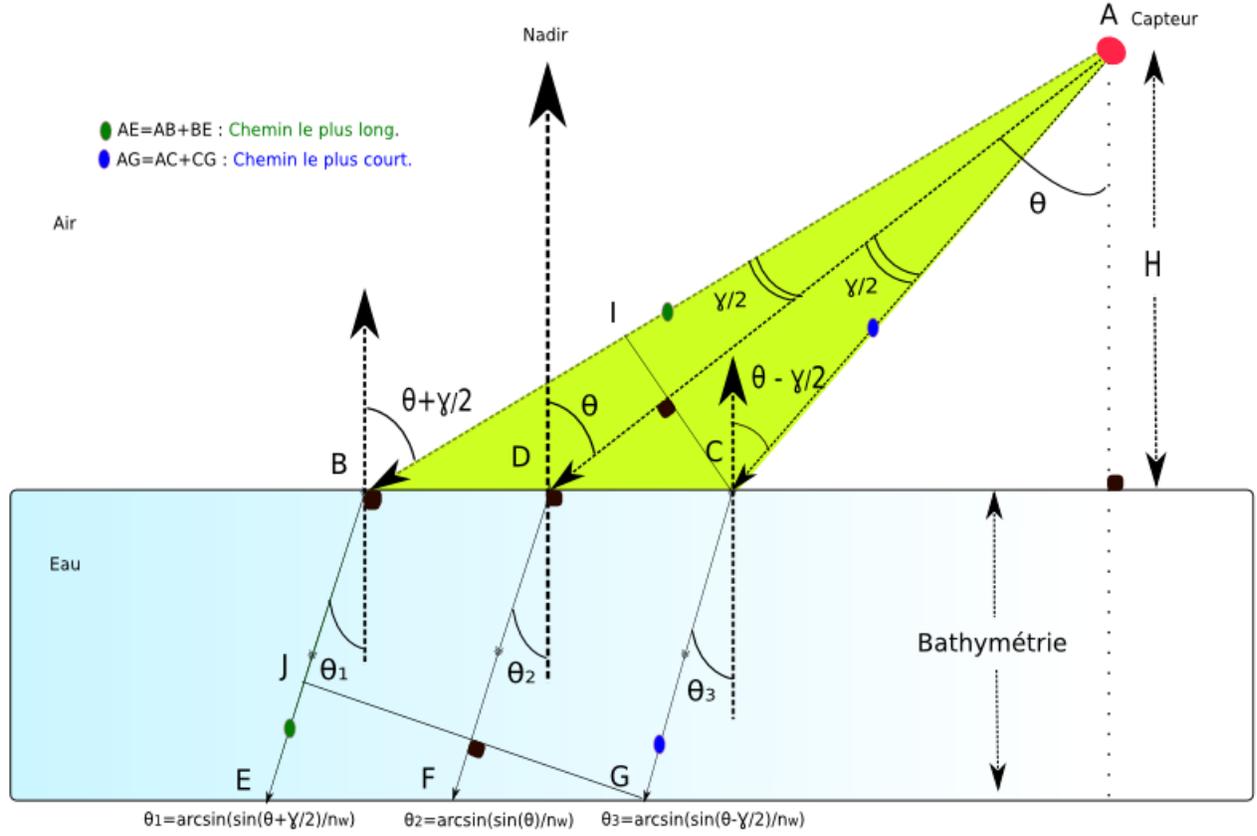


FIGURE 4.5 – Etirement du pulse ("Pulse stretching")

La différence de temps Δt_s aller/retour entre le capteur et la surface de l'eau entre le chemin le plus long AB (long path) et le chemin le plus court AC (short path) que parcourt le laser peut s'écrire comme suit :

$$\Delta t_s = \frac{2H}{c} \left[\frac{1}{\cos(\theta + \frac{\gamma}{2})} - \frac{1}{\cos(\theta - \frac{\gamma}{2})} \right] \quad (4.7)$$

Δt_s est la différence de temps pour que le laser parcoure 2 fois la distance BI (Voir figure 4.5).

La différence de temps Δt_b aller/retour entre la surface de l'eau et le fond de l'eau entre le chemin le plus long BE (long path) et le chemin le plus court CG (short path) peut s'écrire comme suit :

$$\Delta t_b = \frac{2Z}{\left(\frac{c}{n_w}\right)} \left[\frac{1}{\cos(\theta_1)} - \frac{1}{\cos(\theta_3)} \right] \quad (4.8)$$

Δt_b est le temps pour parcourir 2 fois la distance JE (Voir figure 4.5).

- n_w est l'indice de réfraction de l'eau.
- c est la vitesse de la lumière dans l'air.
- θ_1 et θ_3 sont les angles de réfractions illustrés sur la figure 4.5.

L'effet de la différence de temps Δt_s permet de modifier la largeur de la forme de l'onde par un effet d'étirement (stretching) τ_s à la surface de l'eau. Selon Guenther (1985) [23] pour :

$$\Delta t_s < 2 T_0 \Rightarrow \tau_s = \epsilon \Delta t_s \text{ avec } 0.1 < \epsilon < 0.35. \quad (4.9)$$

$$\Delta t_s > 2 T_0 \Rightarrow \tau_s = T_0 \left(\frac{0.5 \Delta t_s}{T_0} - 0.4 \right) \quad (4.10)$$

Dans l'air, la largeur à mi-hauteur du pulse est de T_0 . A la surface de l'eau, la largeur à mi-hauteur du pulse s'étire et devient :

$$\boxed{T_1 = T_0 + \tau_s} \quad (4.11)$$

L'effet de la différence de temps Δt_b permet de modifier une seconde fois la largeur à mi-hauteur du pulse par un effet d'étirement (stretching) τ_b au fond de l'eau. Selon Guenther (1985) [23] pour :

$$\Delta t_b < 2 T_1 \Rightarrow \tau_b = \epsilon \Delta t_b \text{ avec } 0.1 < \epsilon < 0.35. \quad (4.12)$$

$$\Delta t_b > 2 T_1 \Rightarrow \tau_b = T_1 \left(\frac{0.5 \Delta t_b}{T_1} - 0.4 \right) \quad (4.13)$$

L'expression de la largeur à mi-hauteur du pulse au fond de l'eau s'écrit :

$$\boxed{T_2 = T_1 + \tau_b = T_0 + \tau_s + \tau_b} \quad (4.14)$$

Dans notre situation, le pulse est étiré plusieurs fois au cours de son trajet :

- Une première fois à la surface de l'eau dû fait de la différence de temps Δt_s .
- Une seconde fois au fond de l'eau à cause de la différence de temps Δt_b .

4 Puissance rétrodiffusée par la surface de l'eau $P_s(t)$ détectée par le télescope

La forme de d'onde retournée par la surface de l'eau reçue par le détecteur d'après Guenther 1985 [23] est donnée par la relation $P_s(t) = w(t, t_s, \sigma_1) * P_s$ avec :

$$w(t, t_s, \sigma_1) = \frac{2P_e}{\sqrt{2\pi}} e^{-\frac{(t-t_s)^2}{2\sigma_1^2}} \quad (4.15)$$

$$P_s = \frac{P_e T_{atm}^2 A_r \eta_e \eta_r L_s \cos^2(\theta)}{(\pi H^2)} \quad (4.16)$$

- $w(t, t_s, \sigma_1)$ est la forme de l'onde émise par le LiDAR et ayant subi un stretching à la surface de l'eau.
- t_s (s) est le temps aller/retour entre le capteur et la surface de l'eau : $t_s = \frac{2H}{c \cos(\theta)}$.
- σ_1 (s) est l'écart-type du pulse de la gaussienne à la surface de l'eau (section 3).
- P_s (W) est la réponse impulsionnelle à la surface de l'eau.
- P_e (W) est la puissance émise par le laser (section 3).
- T_{atm} (s.u) est la transmittance atmosphérique.
- A_r (m^2) est la surface du capteur.
- η_r (s.u) est l'efficacité optique du récepteur.
- η_e (s.u) est l'efficacité optique du laser.
- L_s (s.u) est la réflectance à l'interface air/eau (facteur de perte) (section 4).
- θ (rad) est l'angle d'incidence entre le rayon incident et la normale à la surface de l'eau (section 4).
- H (m) est l'altitude de vol.
- $P_s(t)$ (W) est la forme de l'onde retournée par la surface de l'eau.

Angle d'incidence θ

Nous définissons l'angle d'incidence θ (en rad), l'angle entre le rayon incident et la normale à la surface de l'eau.

$$\theta = \theta_{nadir} - \theta_s \quad (4.17)$$

- θ_{nadir} (rad) est l'angle entre le rayon incident et le nadir.
- θ_s (rad) est l'angle de la pente à la surface de l'eau.

Les figure 4.6 et 5.2 illustrent l'angle d'incidence selon qu'il y ait ou non une pente à la surface de l'eau.

$\theta_s \neq 0$

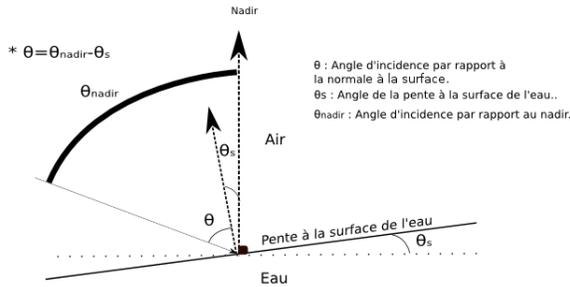


FIGURE 4.6 – Avec une pente à la surface de l'eau.

* $\theta_s = 0 \Rightarrow \theta = \theta_{\text{nadir}}$

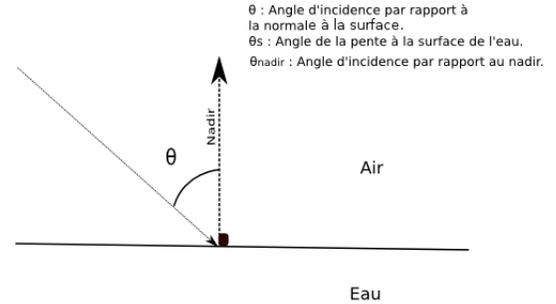


FIGURE 4.7 – Sans pente à la surface de l'eau.

Facteur de perte L_s

Deux méthodes sont décrites dans la littérature pour définir le **facteur de perte** L_s :

- A) La **première méthode** suppose une surface géométrique de l'eau selon le modèle de **Cook, Torrance et Sparrow (1982)** [15] qui représente la surface comme une multitude de microfacettes en V. Le calcul du facteur de perte dans cette méthode consiste à calculer la BRDF de la surface de l'eau diffuse constituée par des microfacettes spéculaires.
- B) La **deuxième méthode** est celle de **Guenther (1985)** [23]. Le facteur de perte est relié à la vitesse du vent w . Cette formule calcule l'effet du vent à la surface de l'eau en fonction de la vitesse du vent. Guenther a également pris en compte la distribution des pentes des vagues en fonction de l'angle d'incidence θ et de la direction du vent suivant la formule de **Cox et Munk (1954)**.

A) Modèle de Cook, Torrance et Sparrow

Le modèle physique de Cook-Torrance et Sparrow [15] repose sur une discrétisation de la surface de l'eau en éléments finis (microfacettes) qui représente la géométrie de la surface de l'eau. Ce modèle est décrit par l'équation suivante :

$$L_s = \frac{1 - k_s}{\pi} + \frac{k_s D O F_r}{\pi \cos^2(\theta)} \quad (4.18)$$

où :

- F_r (s.u) est la fonction de Fresnel (section A.1).
- O (s.u) est le facteur d'atténuation décrivant l'auto-ombrage des facettes les unes par rapport aux autres (section A.2).
- D (s.u) est la fonction de distribution des microfacettes. Elle modélise la rugosité de surface (section A.3).
- k_s (s.u) est le coefficient de réflexion spéculaire.
- θ (rad) est l'angle d'incidence entre le rayon incident et la normale à la surface de l'eau (section 4).

A.1 - Fonction de Fresnel

La fonction de Fresnel F_r dépend de l'angle d'incidence θ . Si $\theta = 0$, l'expression de la fonction de Fresnel s'écrit :

$$F_r = \left(\frac{n_w - 1}{n_w + 1}\right)^2 \tag{4.19}$$

Sinon si $\theta \neq 0$ alors le facteur de Fresnel s'écrit :

$$F_r = \frac{1}{2} \left(\frac{\sin(\theta - \theta_t)}{\sin(\theta + \theta_t)}\right)^2 + \left(\frac{\tan(\theta - \theta_t)}{\tan(\theta + \theta_t)}\right)^2 \tag{4.20}$$

- θ_t (rad) est l'angle de réfraction à l'interface air/eau (section A.1.1).

A.1.1 - Angle de réfraction θ_t

θ_t (rad) est l'angle de réfraction à l'interface air/eau (Figure 4.8). La relation de Snell-Descartes nous indique que :

$$n_{air} \sin(\theta) = n_w \sin(\theta_t) \tag{4.21}$$

$$\sin(\theta_t) = \frac{n_{air} \sin(\theta)}{n_w} \tag{4.22}$$

Avec $n_{air} = 1$, on obtient :

$$\theta_t = \arcsin\left(\frac{\sin(\theta)}{n_w}\right) \tag{4.23}$$

- θ (rad) est l'angle d'incidence à l'interface air/eau entre le rayon incident et la normale à la surface de l'eau (section 4).

- n_w (s.u) est l'indice de réfraction de l'eau.

- n_{air} (s.u) est l'indice de réfraction de l'air.

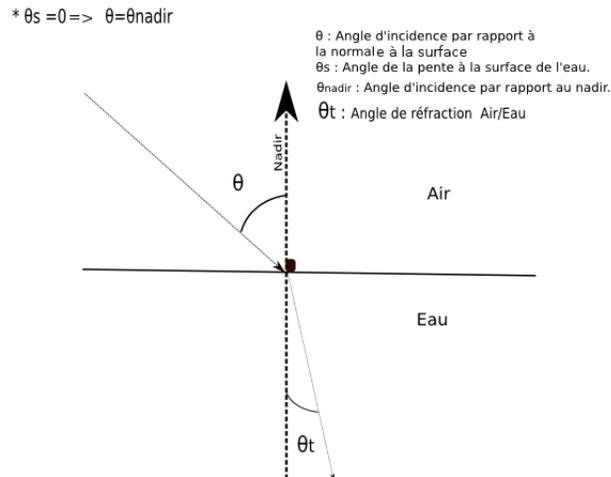


FIGURE 4.8 - θ_t Angle de réfraction air/eau

A.2 - Facteur d'atténuation

L'expression du facteur d'atténuation³ O (Figure 4.9) est obtenue en choisissant la plus petite valeur de ces deux quantités :

$$O = \min(1, 2\cos(\theta)) \quad (4.24)$$

– θ (rad) est l'angle d'incidence entre le rayon incident et la normale à la surface de l'eau.

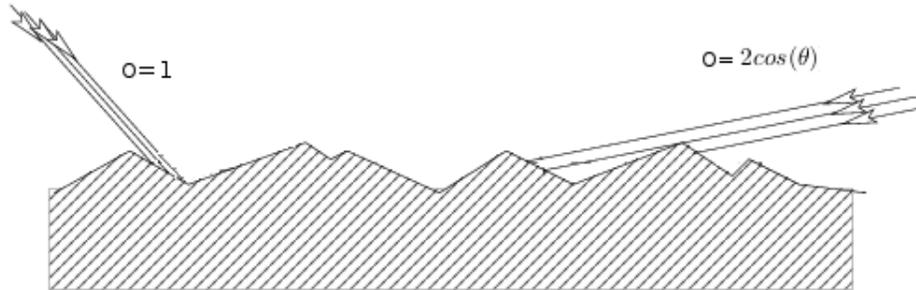


FIGURE 4.9 – L'émetteur et le récepteur du LiDAR sont dans la même direction. La lumière émise ou réfléchié peut-être bloquée ou non par une autre microfacette

A.3 - Fonction de distribution des microfacettes

Le modèle statistique de répartition des microfacettes généralement utilisé est celui issu de Beckmann :

$$D = \frac{1}{r^2 \cos^4(\theta)} \exp\left(-\left(\frac{\tan(\theta)}{r}\right)^2\right) \quad (4.25)$$

– D (s.u) représente la fonction de distribution des microfacettes. Elle modélise **la rugosité de la surface de l'eau**. Dans ce modèle, on considère que la surface de l'eau est constituée d'une multitude de microfacettes (Figure 4.10). Une microfacette est un réflecteur parfait construit par l'assemblage d'un couple symétrique de facettes en forme de V. Chaque microfacette est orientée aléatoirement dans différentes directions suivant une loi de probabilité D .

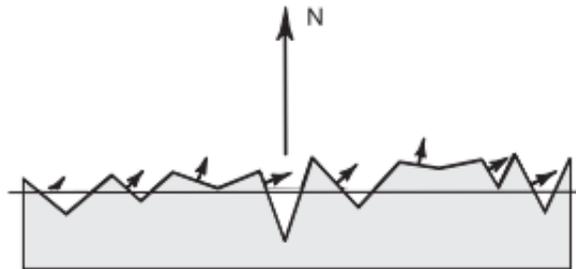


FIGURE 4.10 – Modèle statistique de la surface de l'eau (Référence [80])

3. Pour plus d'informations sur la démonstration voir [79]

- θ (s.u) est l'angle d'incidence entre le rayon incident et la normale à la surface de l'eau.
- r (s.u) (Figure 4.11) est l'écart-type de la pente des microfacettes (ou coefficient de rugosité).

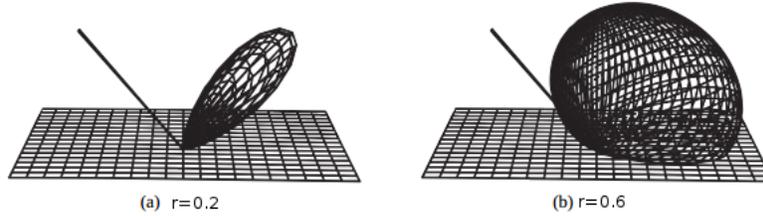


FIGURE 4.11 – Distribution de Beckmann avec (a) $r=0.2$ et (b) $r=0.6$ (Référence [80])

B) Modèle de Guenther et Cox-Munk

La puissance reçue par la surface de l'eau dépend de l'effet du vent et de l'angle d'incidence du faisceau laser. Elle est considérée comme un facteur de perte qui varie entre 10^{-1} et 10^{-7} . Selon Guenther (1985) [23], le facteur de perte L_s peut s'écrire :

$$L_s(\theta, w) = \delta_s(w) N_d(\theta, w) \cos^2(\theta) \quad (4.26)$$

- $N_d(\theta, w)$ (s.u) est l'état normalisé de la distribution de la pente des vagues (section B.1).
- $\delta_s(w)$ (s.u) est la perturbation de la surface de l'eau en fonction du vent (section B.2).
- θ (rad) est l'angle entre le rayon incident et la normale à la surface de l'eau (section 4).

B.1 - Etat normalisé de la distribution de la pente des vagues $N_d(\theta, w)$

Le paramètre $N_d(\theta, w)$ est l'état normalisé de la distribution de la pente des vagues désigné aussi par le facteur de correction de l'albédo. Il est exprimé par Cox-Munk (1954) [16] en fonction de la vitesse du vent dans une direction \mathbf{d} donnée et de l'angle d'incidence θ . Les pentes des vagues varient suivant une distribution gaussienne en fonction de la direction \mathbf{d} du vent et de l'angle d'incidence :

$$N_d = \exp\left(-\frac{\theta^2}{2\sigma_d^2}\right) \quad (4.27)$$

σ_d (s.u) est l'écart-type des pentes des vagues.

L'indice \mathbf{d} représente la direction du vent qui peut-être :

- **u** pour une **direction up/down du vent** par rapport à la trajectoire du capteur.
- **c** pour une **direction cross du vent** par rapport à la trajectoire du capteur.

D'après Cox-Munk (1954) [16] :

- si on a une **direction up/down du vent** alors l'écart type des pentes des vagues qui est le rms (root mean square) des pentes des vagues est défini par :

$$\sigma_u = 0.00158w \quad (4.28)$$

- si on a une **direction cross du vent** alors :

$$\sigma_c = 0.003 + 0.00096w \quad (4.29)$$

B.2 -Expression de la perturbation en fonction du vent $\delta_s(w)$

Si la vitesse du vent w est comprise entre 3 et 17 noeuds, l'expression de la perturbation en fonction du vent w au nadir (en %) s'écrit :

$$\delta_s(w) = 214 - \left(\frac{w - 4.8}{12.2}\right)^{2.06} + 37 \quad \text{avec } 3 \leq w \leq 17 \quad (4.30)$$

Et si la vitesse du vent w est supérieure à 17 noeuds, l'expression de la perturbation de la surface en fonction du vent w au nadir (en %) est :

$$\delta_s(w) = 4.26 \exp(-0.14w) \quad \text{avec } w > 17 \quad (4.31)$$

La perturbation de la surface $\delta_S(w)$ est donnée par Petri (1977) [43].

5 Puissance rétrodiffusée $P_c(t)$ par la colonne d'eau détectée par le télescope

Pour pouvoir calculer la puissance rétrodiffusée par la colonne d'eau détectée par le télescope, on modélise en amont la colonne d'eau comme **une multitude de couches** d'eau (Figure 4.12).

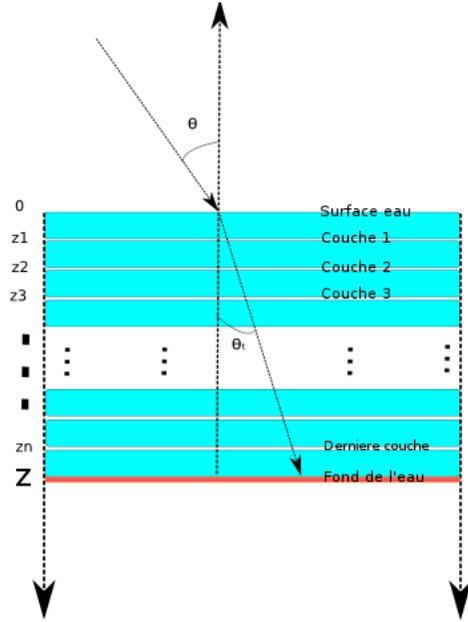


FIGURE 4.12 – Les profondeurs des couches de la colonne d'eau.

De manière, ensuite, à calculer la réponse impulsionnelle de chaque couche de la colonne d'eau. D'après Guenther [23], on a :

$$P_c(z_b) = \frac{P_e T_{atm}^2 A_r \eta_e \eta_r F (1 - L_s)^2 \beta(\phi)}{\frac{n_w * H + z_b}{\cos^2(\theta)}} \exp\left(-\frac{2K z_b}{\cos(\theta_t)}\right) \quad (4.32)$$

- P_e (W) est la puissance émise par le laser
- T_{atm} (s.u) est la transmittance atmosphérique
- A_r (m^2) est la surface du capteur
- η_e (s.u) est l'efficacité optique du laser
- η_r (s.u) est l'efficacité optique du récepteur
- F (s.u) est le facteur de perte dû au FOV qui a des valeurs entre 0 et 1.
- L_s (s.u) est le facteur de perte
- K (m^{-1}) est le coefficient d'atténuation total de la colonne d'eau
- $\beta(\phi)$ (m^{-1}) est la fonction du volume diffusé (volume scattering function) pour un angle de diffusion (Petzold (1972)).
- θ_t (rad) est l'angle d'incidence dans l'eau
- n_w (s.u) est l'indice de réfraction de l'eau
- H (m) est l'altitude du capteur

$P_c(z_b)$ est un **vecteur** qui s'écrit comme suit :

$$P_c(z_b) = \begin{pmatrix} P_c(z_1) \\ P_c(z_2) \\ P_c(z_3) \\ \cdot \\ \cdot \\ \cdot \\ P_c(z_n) \end{pmatrix}$$

z_b est un **vecteur** représentant les différentes profondeurs des couches dans la colonne d'eau (Figure 4.12).

$$z_b = \frac{(t_c - t_s) \left(\frac{c}{n_w}\right) \cos(\theta_t)}{2} \quad \text{avec } t_c > t_s \quad (4.33)$$

$$z_b = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ \cdot \\ \cdot \\ \cdot \\ z_n \end{pmatrix}$$

avec z_1, z_2, \dots, z_n respectivement la profondeur de la première, la deuxième, ... et la dernière couche de la colonne d'eau (figure 4.12).

$$t_c = \begin{pmatrix} t_s + t_1 \\ t_s + t_2 \\ t_s + t_3 \\ \cdot \\ \cdot \\ \cdot \\ t_s + t_n \end{pmatrix}$$

avec t_1, t_2, \dots, t_n respectivement le temps aller/retour que met le laser pour aller de la surface de l'eau à la couche z_1, z_2, \dots, z_n .

La forme de l'onde retournée par la colonne d'eau (figure 4.13) est :

$$\boxed{P_c(t) = w(t, t_c, \sigma_1) \otimes P_c(z_b)} \quad (4.34)$$

$$w(t, t_c, \sigma_1) = \frac{2P_e}{\sqrt{2\pi}} \exp\left(-\frac{(t - t_s)^2}{2 * \sigma_1^2}\right) \quad (4.35)$$

avec \otimes désignant le produit de convolution (voir figure 4.13).

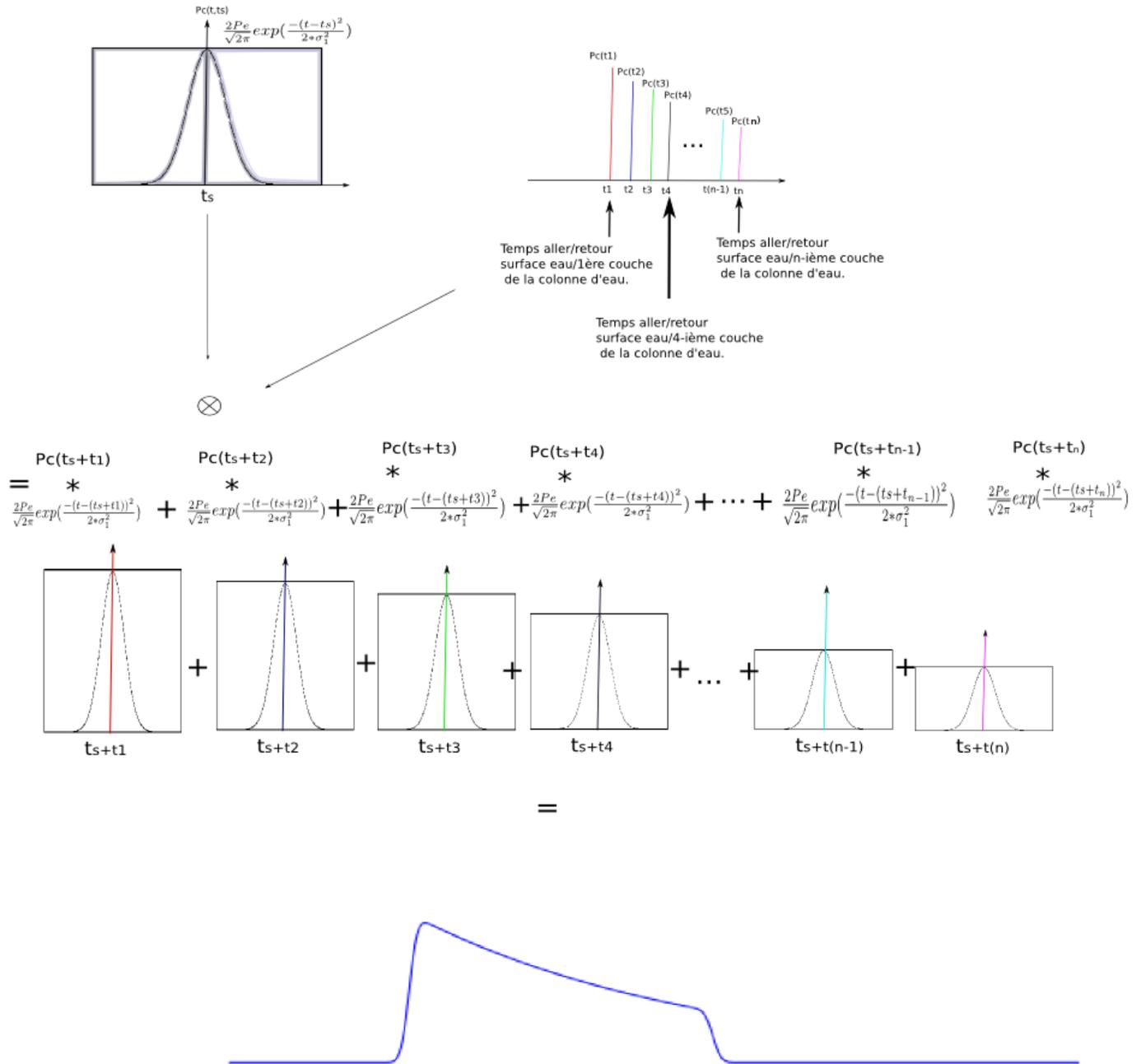


FIGURE 4.13 – Produit de convolution. Impulsion gaussienne émise par le LiDAR convoluée avec la réponse impulsionnelle de chaque couche de la colonne d'eau

6 Puissance rétrodiffusée par le fond de l'eau $P_b(t)$ détectée par le télescope

La forme d'onde retournée au capteur par le fond de l'eau est donné par Guenther [23] par la relation suivante :

$$\boxed{P_b(t) = P_b(Z) * w(t, t_b, \sigma_2)} \quad (4.36)$$

avec :

$$P_b(Z) = \frac{P_e T_{atm}^2 A_r \eta_e \eta_r F R_b (1 - L_s)^2 \exp\left(-\frac{2kZ}{\cos(\theta_t)}\right)}{\pi \left(\frac{n_w H + Z}{\cos(\theta)}\right)^2} \quad (4.37)$$

$$w(t, t_b, \sigma_2) = \frac{2P_e}{\sqrt{2\pi}} \exp\left(-\frac{(t - t_b)^2}{2\sigma_2^2}\right) \quad (4.38)$$

- t_b (s) est le temps aller/retour entre le détecteur et le fond de l'eau : $t_b = t_s + \frac{2Z}{\frac{c}{n_w} \cos(\theta_t)}$.
- σ_2 (s.u) est l'écart-type de la gaussienne au fond de l'eau (section 3).
- T_{atm} (s.u) est la transmittance atmosphérique.
- $w(t, t_b, \sigma_2)$ est la forme de l'onde émise par le Lidar et ayant subit un stretching à la surface de l'eau.
- $P_b(Z)$ (W) est la réponse impulsionnelle du fond de l'eau.
- P_e (W) est la puissance émise par le laser (section 3).
- A_r (m^2) est la surface du capteur.
- η_r (s.u) est l'efficacité optique du receptr.
- η_e (s.u) est l'efficacité optique du laser.
- L_s (s.u) est la réflectance à l'interface air/eau (facteur de perte).
- θ_t (rad) est l'angle de réfraction à l'interface air/eau.
- F (s.u) est le facteur de perte dû au FOV qui a des valeurs entre 0 et 1.
- R_b est l'albédo du fond de l'eau
- Z (m) est la bathymétrie
- $P_b(t)$ (W) est la forme de l'onde retournée par le fond de l'eau.

7 Sources de bruit détectées par le télescope

Les sources de bruit sont très variées suivant les systèmes de télédétection utilisés. Pour ce qui est d'une mesure LiDAR, il existe deux sources de bruit principales celle due au rayonnement solaire et celle générée par le détecteur.

7.1 Bruit dû à la radiation solaire détecté par le télescope $P_{bg}(t)$

L'équation pour la radiation solaire est donnée par Marshall and Burk (1986) [62] :

$$P_{bg} = I_s T_{atm} A_r (1 - \gamma_r^2) \pi \frac{FOV^2}{4} \Delta_B \eta_r \eta_\lambda \quad (4.39)$$

$$P_{bg}(t) = P_{bg} * \underbrace{y}_{= \text{Bruit blanc gaussien}} \quad (4.40)$$

avec $y \rightarrow N(0, 1)$ vecteur de n valeurs aléatoires représentant le bruit blanc gaussien.

- I_s (W/m^{-2}) est la valeur de l'irradiance du soleil .
- T_{atm} (s.u) est la transmittance atmosphérique.
- A_r (m^2) est la surface du capteur.
- γ_r (s.u) est le ratio d'obscurcissement reçu par le télescope.
- FOV⁴ (rad) est le champ de vue du télescope.
- Δ_B (m) est la largeur de bande du filtre optique du détecteur.
- η_r (s.u) est l'efficacité optique du récepteur.
- η_λ (s.u) est la transmission de bande du filtre.

7.2 Bruit du détecteur détecté par le télescope $P_N(t)$

D'après Chisheng Wang et al. [65] , on calcule le bruit du détecteur par :

$$\sigma_N = \frac{Max(P_s(t) + P_c(t) + P_b(t))}{PSNR} \quad (4.41)$$

où :

PSNR (s.u) est le signal rapport sur bruit (Signal to Noise Ratio).

$P_s(t)$ (W) est la puissance reçue sur le détecteur de la surface de l'eau.

$P_c(t)$ (W) est la puissance reçue sur le détecteur de la colonne l'eau.

$P_b(t)$ (W) est la puissance reçue sur le détecteur du fond de l'eau.

Le bruit du détecteur est assimilé à une loi de distribution normale de moyenne 0 et de variance σ_N :

$$\boxed{P_N(t) = N(0, \sigma_N)} \quad (4.42)$$

8 Exemple d'une forme d'onde simulée avec le simulateur Wa-LID 2.0

La figure 4.14 illustre une forme d'onde LiDAR bathymétrique générée avec le simulateur Wa-LID 2.0. On peut également y voir les 5 puissances qui composent une forme d'onde et la puissance émise pour une configuration d'un système LiDAR bathymétrique.

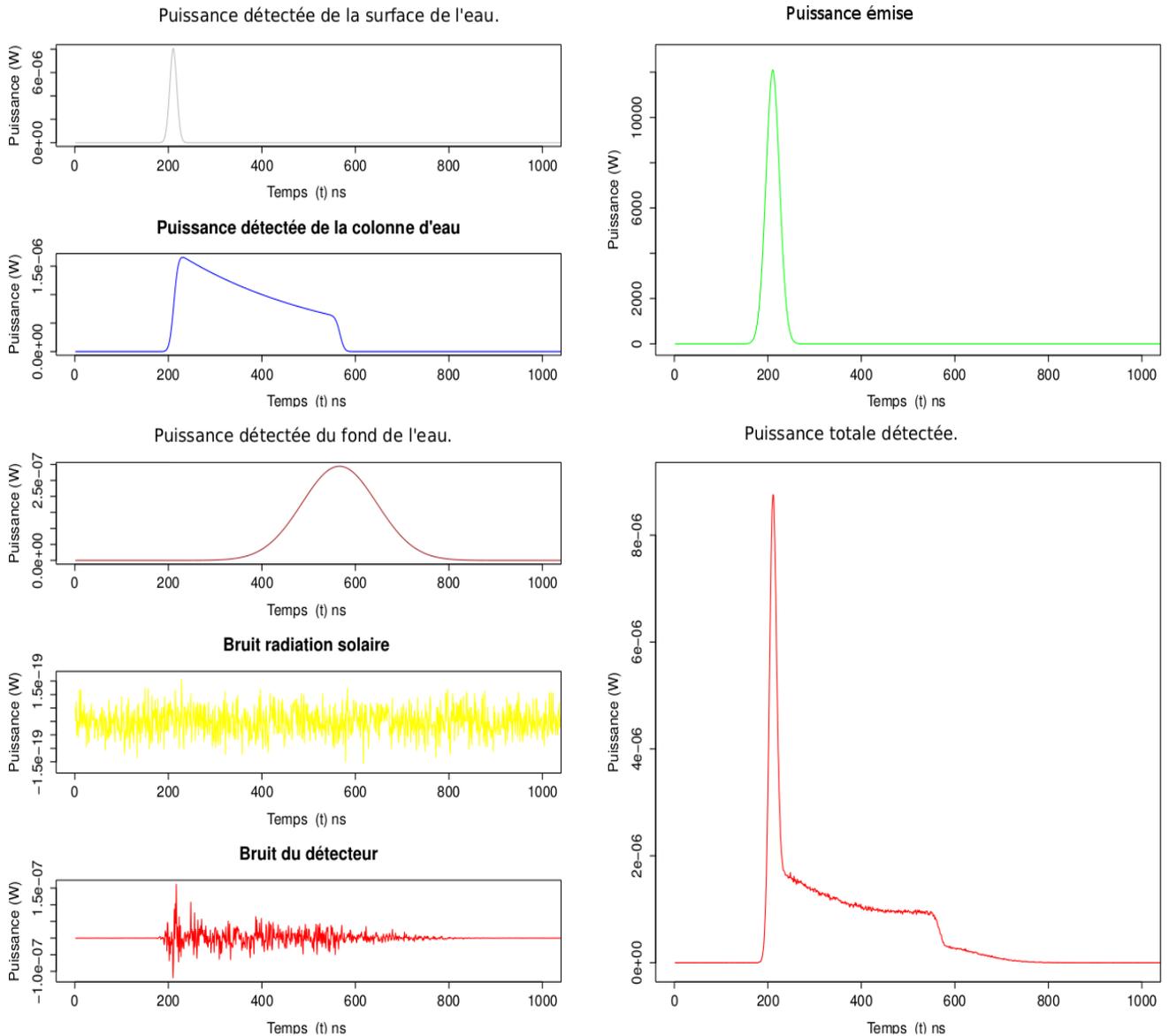


FIGURE 4.14 – Exemple d'une forme d'onde simulée avec le simulateur Wa-LID 2.0

9 Intérêts du simulateur bathymétrique Wa-LID 2.0

Le but du simulateur Wa-LID 2.0 est de pouvoir générer un grand nombre de formes d'onde correspondant à de nombreuses configurations instrumentales et environnementales (très difficile à obtenir sans simulateur) et ainsi développer des méthodes d'inversion des formes d'onde pour l'estimation de la bathymétrie. L'utilisateur (Figure 4.15) peut choisir un certain nombre de paramètres tels que :

- les **paramètres instrumentaux** du LiDAR (longueur d'onde, FOV⁵, surface du capteur,...)
- les **paramètres environnementaux** (Coefficient d'absorption de l'eau, concentration de substances jaunes, concentration de phytoplancton...)
- **la profondeur de l'eau ...**

On peut donc générer un grand nombre de formes d'onde LiDAR bathymétriques correspondant à différentes configurations du milieu et d'instrument et **estimer** à partir de celles-ci **la bathymétrie**. J'ai ainsi généré **deux bases de données** de 780 formes d'onde avec une eau claire et une eau turbide.

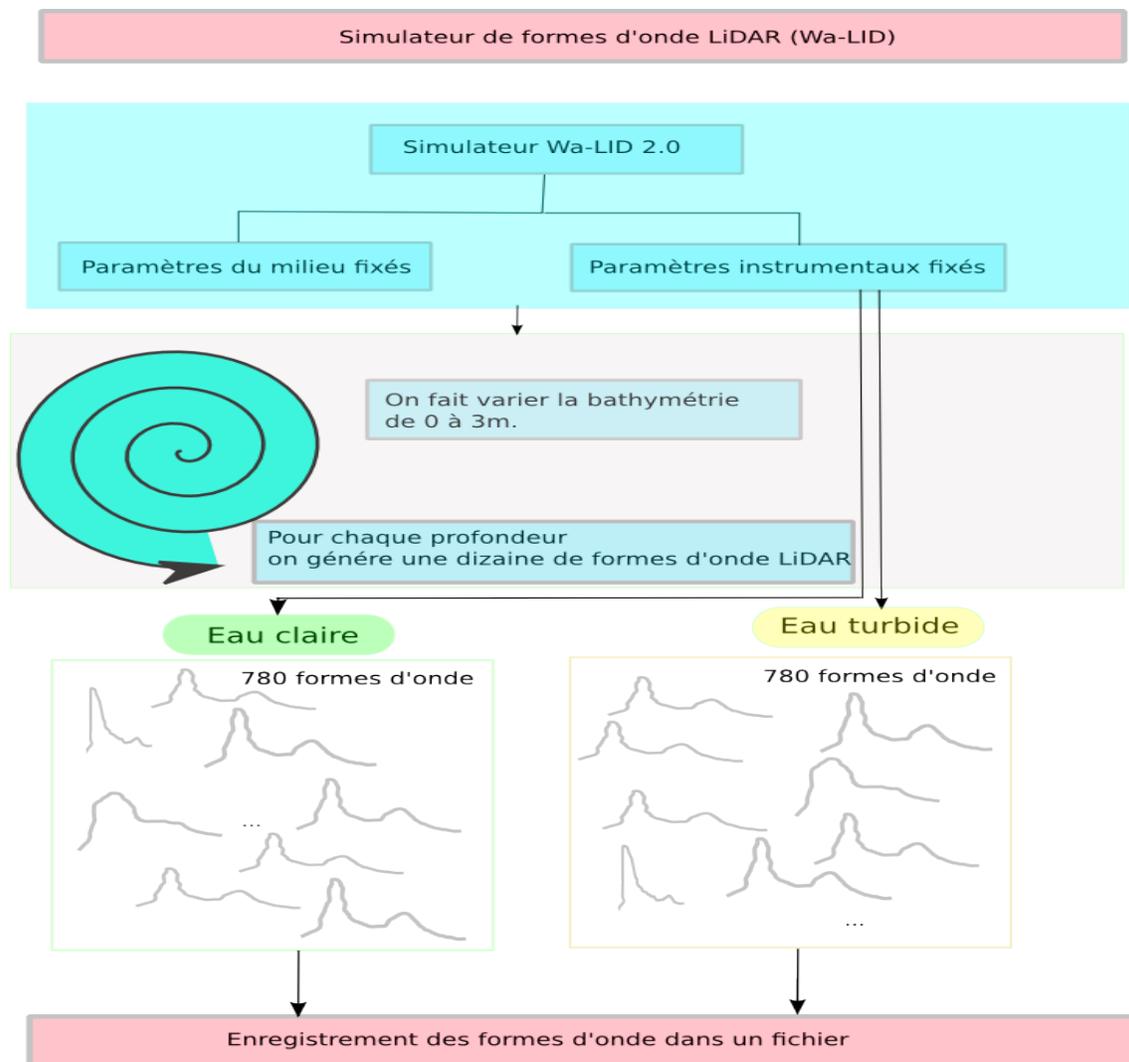


FIGURE 4.15 – Simulation de deux bases de données de 780 formes d'onde LiDAR

5. Field Of View

Deuxième partie

Les données LiDAR utilisées

Chapitre 5

Description des données

Sommaire

| | | |
|----------|---|-----------|
| 1 | Formes d'onde bathymétriques | 46 |
| 1.1 | Formes d'onde réelles : Site du Golfe du Morbihan | 46 |
| 1.2 | Formes d'onde simulées pour deux configurations du milieu | 48 |
| 2 | Formes d'onde altimétriques | 49 |
| 2.1 | Formes d'onde réelles : Forêt du Brésil | 49 |

Résumé du chapitre :

Dans ce chapitre, je présente **cinq bases de données de formes d'onde LiDAR**. Grâce au simulateur Wa-LID 2.0, j'ai généré deux bases de données de formes d'onde LiDAR bathymétrique avec une gamme de profondeurs comprises entre **0.5 et 3 mètres**. Les 3 bases de données de formes d'onde réelles restantes ont été acquises sur **le site du Golfe du Morbihan** pour des profondeurs allant de **0 à 3 mètres** et sur **le site du Brésil** pour des hauteurs d'arbres allant de **0 à 35 mètres**

1 Formes d'onde bathymétriques

1.1 Formes d'onde réelles : Site du Golfe du Morbihan

Les données LiDAR bathymétriques ont été acquises en Juin 2005 sur le Golfe du Morbihan (Figure 5.1) par le Service Hydrographique et Océanographique de la Marine (SHOM) à l'aide de deux détecteurs de type : photodiode à avalanche (APD) (section 1.1.1) et un tube photomultiplicateur (PMT) (section 1.1.2)

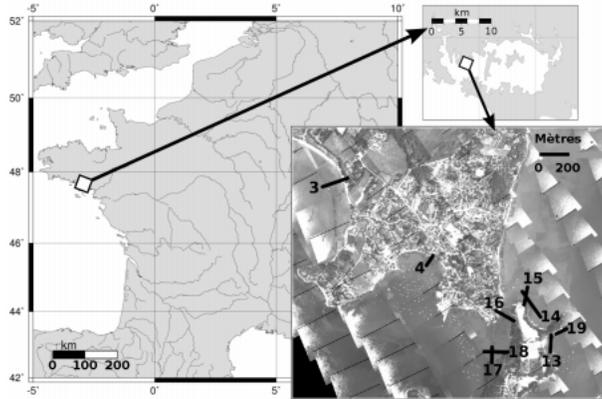


FIGURE 5.1 – Le Golfe du Morbihan en Bretagne sud (Référence : Tristan et al).

Je dispose donc de **deux jeux de données de formes d'onde réelles bathymétriques** acquises avec deux types de capteurs. Les sections 1.1.1 et 1.1.2 présentent ces deux jeux de données.

1.1.1 Le capteur Avalanche PhotoDiode (APD)

La base de données du Golfe du Morbihan acquise avec le capteur APD comprend :

- **189 formes d'onde**
- et une **bathymétrie mesurée** sur le terrain pour **chaque forme d'onde**.

Une **forme d'onde** est le résultat de l'enregistrement de **200 points** de mesure acquis par un LiDAR bathymétrique. Un point correspond à la puissance reçue par le lidar à un instant donné. Chaque point est enregistré à un intervalle de temps régulier de 1 ns. Ci-dessous, le tableau 5.1 résume la structure du jeu de données relatif au site du Golfe du Morbihan avec le capteur APD :

| - | Bathymétrie mesurée | $P(t_1)$ | $P(t_2)$ | ... | ... | $P(t_{200})$ |
|--------------------|---------------------|----------|----------|-----|-----|--------------|
| Forme d'onde n°1 | 0.9288 | 8 | 7 | .. | ... | 9 |
| Forme d'onde n°2 | 0.6158 | 8 | 9 | ... | ... | 8 |
| ... | ... | ... | ... | ... | ... | ... |
| Forme d'onde n°188 | 2.1400 | 9 | 9 | ... | ... | 5 |
| Forme d'onde n°189 | 2.3797 | 7 | 7 | ... | ... | 8 |

TABLE 5.1 – Tableau de 189 formes d'onde enregistrées par un LiDAR bathymétrique avec le capteur APD.

Dans la figure 5.2, je présente, à titre d'exemple, les formes d'onde n°1 et n°166 contenues dans la base de données.

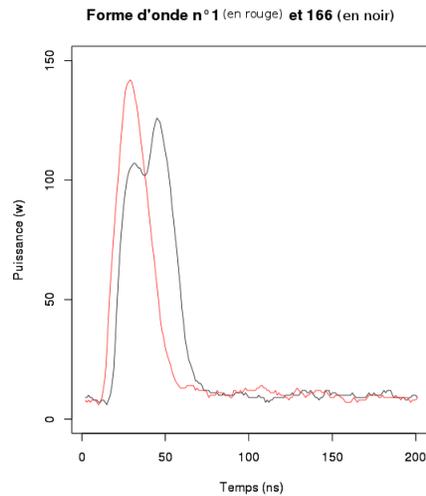


FIGURE 5.2 – Représentation de deux formes d’onde contenues dans la base de données du Golfe du Morbihan. La bathymétrie mesurée sur le terrain pour les forme d’onde n°1 et n°166 est respectivement de 92,88 cm et 2.4 m.

1.1.2 Le capteur PhotoMultipliers Tubes (PMT)

La base de données du Golfe du Morbihan acquise avec le capteur PMT comprend :

- **189 formes d’onde**
- et une **bathymétrie mesurée** sur le terrain pour **chaque forme d’onde**.

Une **forme d’onde** est le résultat de l’enregistrement de **500 points** de mesure acquis par un LiDAR bathymétrique. Un point correspond à la puissance reçue par le LiDAR à un instant donné. Chaque point est enregistré à un intervalle de temps régulier de 1 ns.

1.2 Formes d'onde simulées pour deux configurations du milieu

Avec le simulateur Wa-LID 2.0, j'ai généré deux jeux de données de 780 formes d'onde simulées en faisant varier la bathymétrie¹ de 0.5 à 3 mètres avec un pas de 10 centimètres. Une forme d'onde simulée est constituée de 512 points de mesures à intervalles de temps réguliers de 1 ns. J'ai donc au préalable fixé les paramètres instrumentaux de sorte à pouvoir me placer dans la configuration d'un capteur de type : Satellite EADS (Tableau 5.2).

| Paramètres | Satellite (EADS) |
|--|------------------|
| Wavelength λ (nm) | 532 |
| Altitude H (km) | 500 |
| Laser energy E0 (J) | 43 |
| Pulse duration T_0 (ns) | 3.5 |
| Incidence angle θ (°) | 0 |
| Receiver area Ar (m^2) | 0.63 |
| Field of view FOV (rad) | 9.10^{-5} |
| Responsivity R_λ (A/W) | 0.25 |
| Divergence angle (rad) | 6.10^{-5} |
| Emitted optical efficiency η_e (s.u) | 0.8 |
| Receiver optical efficiency η_r (s.u) | 0.5 |
| Loss factor F (s.u) | 1 |
| Filter's bandwidth Δ_b (nm) | 1 |
| Receiver obturation γ_r (s.u) | 0.1 |
| Electrical bandwidth B (MHz) | 500 |
| Excess noise G_{noise} (s.u) | 3 |
| Dark current Id (A) | 10^{-10} |

TABLE 5.2 – Paramètres instrumentaux

1.2.1 Formes d'onde simulées pour une eau claire

Dans la première simulation, nous avons fixé les paramètres de l'eau de sorte à avoir une eau claire (Tableau 5.3).

| Paramètres du milieu | Valeurs |
|--|---------|
| Pente de surface de l'eau (°) | 0.00019 |
| Réflectance bidirectionnelle spéculaire k_s (s.u) | 0.2 |
| Rugosité de surface r (s.u) | 0.45 |
| Concentration en phytoplancton C_{ph} ($mg\ m^{-3}$) | 0.4 |
| Concentration en substances jaunes a_y (m^{-1}) | 0.01 |
| Concentration en sédiments S ($mg\ l^{-1}$) | 2.6 |
| Albédo de fond Rb (s.u) | 0.1 |

TABLE 5.3 – Paramètres pour une eau claire

1. Pour chaque profondeur nous avons généré plusieurs formes d'onde (environ une dizaine).

1.2.2 Formes d'onde simulées pour une eau plus turbide

Dans la deuxième simulation, j'ai fixé les paramètres de manière à avoir une eau plus turbide (Tableau 5.4).

| Paramètres du milieu | Valeurs |
|--|---------|
| Pente de surface de l'eau ($^{\circ}$) | 0.00019 |
| Réflectance bidirectionnelle spéculaire k_s (s.u) | 0.84 |
| Rugosité de surface r (s.u) | 0.23 |
| Concentration en phytoplancton C_{ph} ($mg\ m^{-3}$) | 24.12 |
| Concentration en substances jaunes a_y (m^{-1}) | 0.001 |
| Concentration en sédiments S ($mg\ l^{-1}$) | 4.49 |
| Albédo de fond R_b (s.u) | 0.17 |

TABLE 5.4 – Paramètres pour une eau plus turbide

2 Formes d'onde altimétriques

2.1 Formes d'onde réelles : Forêt du Brésil

Les données LiDAR altimétrique ont été acquises au Brésil par le capteur LiDAR GLAS² embarqué sur le satellite IceSAT.

La base de données est constituée :

- **305 formes d'onde**
- et d'une **hauteur d'arbre mesurée** sur le terrain pour **chaque forme d'onde** (entre 0 et 31 mètres).

Une **forme d'onde** est le résultat de l'enregistrement de **1000 points** de mesure acquis par un LiDAR altimétrique. Un point correspond à la puissance recue par le LiDAR à un instant donné. Chaque point est enregistré à un intervalle de temps régulier de 1 ns.

La figure 5.3 représente la parcelle d'Eucalyptus du Brésil dont nous souhaitons estimer les hauteurs d'arbres.



FIGURE 5.3 – Photo d'une parcelle d'Eucalyptus en cours de coupe ©Cirad

Chapitre 6

Préparation des données LiDAR

Sommaire

| | | |
|----------|---|-----------|
| 1 | Partie utile W_{ext} d'une forme d'onde LiDAR | 51 |
| 1.1 | Partie utile W_{ext} d'une forme d'onde LiDAR | 51 |
| 1.2 | Intérêts de calculer la partie utile d'une forme d'onde LiDAR | 51 |
| 1.3 | Comment calculer la partie utile d'une forme d'onde LiDAR? | 51 |
| 2 | Données LiDAR après préparation | 54 |

Résumé du chapitre :

Avant de pouvoir appliquer les modèles d'inversion sur des formes d'onde LiDAR, je dois au préalable "préparer les données LiDAR". Pour chaque forme d'onde, on ne va garder que la partie utile du signal qui est située entre la surface et le fond de l'eau pour les formes d'onde bathymétrique et entre le haut de l'arbre et le sol pour les formes d'onde altimétrique avec la contrainte que les formes d'onde issues de ce découpage ont le même nombre de points de mesure. Pour cela, j'introduis la notion de "partie utile d'une forme d'onde" (W_{ext}).

1 Partie utile W_{ext} d'une forme d'onde LiDAR

1.1 Partie utile W_{ext} d'une forme d'onde LiDAR

Pour chaque forme d'onde, je m'intéresse seulement à une partie du signal. C'est la partie de la forme d'onde qui contient le premier et dernier pic (correspondant respectivement au signal rétrodiffusé par la surface et le fond de l'eau dans le cas de la bathymétrie) qui nous permettra d'estimer la bathymétrie. C'est ce que nous appelons "**la partie utile de la forme d'onde**", elle est notée W_{ext} . La figure 6.1 illustre la partie utile d'une forme d'onde.

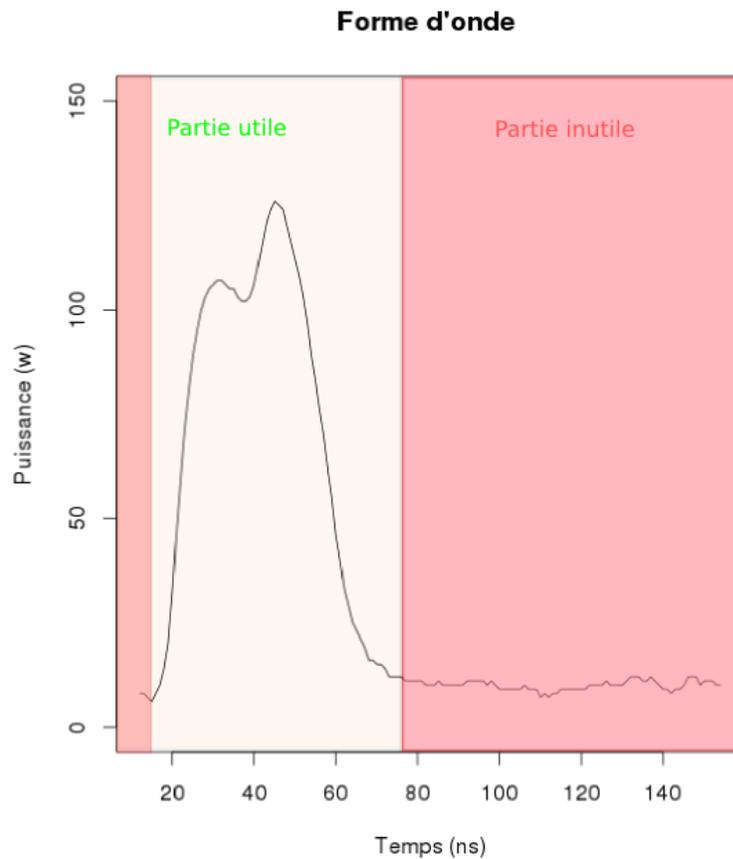


FIGURE 6.1 – Partie utile d'une forme d'onde LiDAR.

1.2 Intérêts de calculer la partie utile d'une forme d'onde LiDAR

Nous calculons la partie utile W_{ext} de chaque forme d'onde car d'une part cela nous permet de garder la partie du signal qui nous renseigne sur le paramètre à estimer (bathymétrie ou hauteur d'arbre) et d'autre part, car c'est une variable qui est corrélée à la bathymétrie. En effet, plus la partie utile est grande et plus la bathymétrie est grande.

1.3 Comment calculer la partie utile d'une forme d'onde LiDAR ?

Les figures 6.2, 6.3, 6.4 et 6.5 illustrent les étapes pour calculer la partie utile du signal d'une forme d'onde.

- **Étape 1** : On sélectionne une partie du signal correspondant au bruit parasite du signal (bruit dû à la radiation solaire, bruit interne du capteur...). Dans la figure 6.2, elle est représentée par un cadran orange (de 80 à 140 ns).

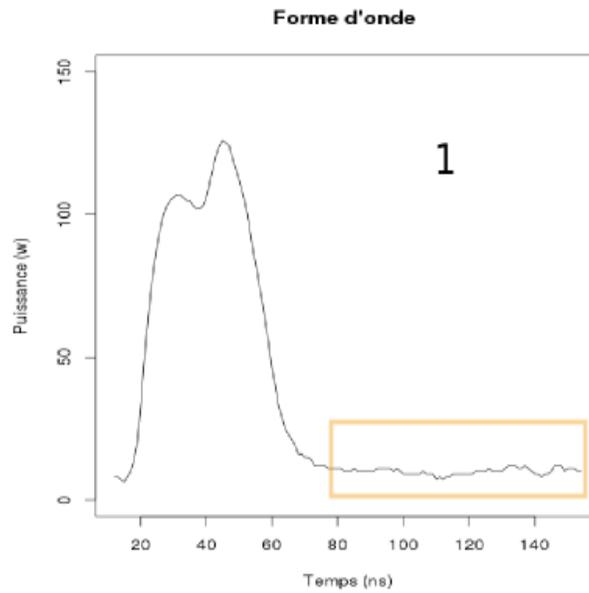


FIGURE 6.2 – Etape 1 : la sélection du bruit

- **Étape 2** : On calcule ensuite la puissance moyenne du bruit dans la partie du signal sélectionnée à l'étape 1. Sur la figure 6.3, elle est représentée par la droite rouge.

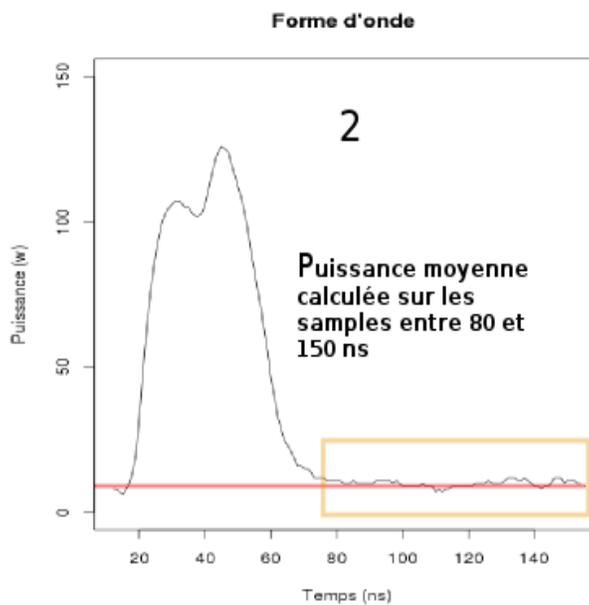


FIGURE 6.3 – Etape 2 : Calcul de la puissance moyenne du bruit : ligne en rouge

- **Etape 3** : On calcule l'écart-type sur la partie du signal correspondant au bruit. Sur la figure 6.4, on trace la droite verte correspondant à la puissance moyenne du bruit (calculée à l'étape 2) + 5 fois l'écart-type, c'est à dire $y = \text{Puissance moyenne} + 5 * \text{ecart-type}$. La droite verte corespond au seuil délimitant le bruit parasite du signal.

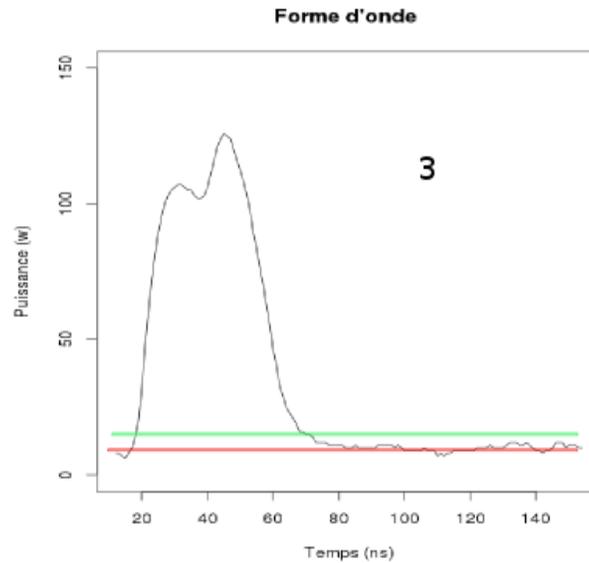


FIGURE 6.4 – Etape 3 : Seuil délimitant le bruit parasite du signal : ligne verte

- **Etape 4** : On récupère les abscisses correspondant aux intersections de la droite verte et de la forme d'onde. L'écart en valeur absolue des deux abscisses corespondent à la partie utile du signal. Sur la figure 6.5, la partie utile du signal commence à $x_d = 19$ ns et finit à $x_f = 73$ ns. La partie utile de cette forme d'onde est $W_{ext} = x_f - x_d$.

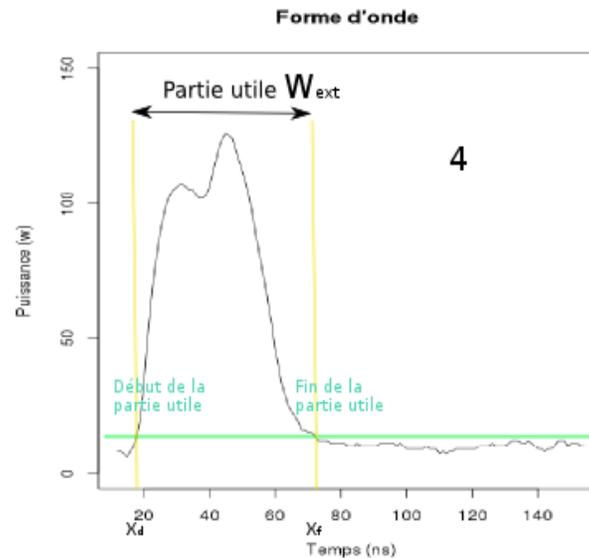


FIGURE 6.5 – Etape 4 : Calcul de la partie utile d'une forme d'onde LiDAR

2 Données LiDAR après préparation

J'ai ensuite préparé mes données en calculant le début x_d^i et la fin x_f^i de la partie utile de chaque forme d'onde (i représentant le numéro de la forme d'onde¹). Pour une base de données de N formes d'onde, je garde la partie du signal entre x_d^i et $x_d^i + \max(W_{ext}^i)$.

| Base de données | Nombre de formes d'onde | Nombre de points de mesure du signal avant préparation | Nombre de points de mesure du signal après préparation |
|------------------------|-------------------------|--|--|
| Golfe Morbihan (APD) | 189 | 200 | 64 |
| Golfe Morbihan (PMT) | 189 | 500 | 128 |
| Simulateur Eau Turbide | 780 | 582 | 64 |
| Simulateur Eau Claire | 780 | 582 | 64 |
| Forêt Brésil | 305 | 1000 | 512 |

TABLE 6.1 – Caractéristiques des bases de données

1. pour une base donnée

Troisième partie

Les différentes méthodes d'inversion des formes d'onde LiDAR

Chapitre 7

Vers de nouvelles méthodes d'inversion de formes d'onde LiDAR

Sommaire

| | | |
|----------|--|-----------|
| 1 | Les méthodes d'inversion des formes d'onde LiDAR rencontrées dans la littérature et leurs limites | 57 |
| 1.1 | Etat de l'art des méthodes d'inversion de formes d'onde LiDAR dans la littérature | 57 |
| 1.2 | Les limites des méthodes d'inversion de formes d'onde actuelles | 58 |
| 2 | Vers de nouvelles méthodes d'inversion de formes d'onde LiDAR | 59 |

Résumé du chapitre : Il s'agit ici d'un **chapitre clé** qui me permet de faire référence aux techniques d'inversion des formes d'onde LiDAR rencontrées dans la littérature. J'introduis, ensuite, l'approche statistique des nouvelles méthodes d'inversion de formes d'onde LiDAR que je vais utiliser.

1 Les méthodes d'inversion des formes d'onde LiDAR rencontrées dans la littérature et leurs limites

1.1 Etat de l'art des méthodes d'inversion de formes d'onde LiDAR dans la littérature

Dans la littérature [17] et [76], on peut distinguer plusieurs types de méthodes d'inversion du signal :

- **les approximations mathématiques** : une forme d'onde est approximée par une fonction mathématique, les paramètres de la fonction pouvant être traduits en variables d'intérêt.
- **les méthodes heuristiques (déterministes ou expertes)** : utilisation successive de règles de décision en vue de reconnaître la surface et le fond tout en rejetant les artefacts créés par le matériel ou l'environnement.
- **la déconvolution** : Ce type de méthode permet de supprimer la composante du signal transmis contenue dans le signal rétrodiffusé. La déconvolution permet de corriger la perte de précision causée par une fonction de réponse du récepteur LiDAR plus longue que la résolution temporelle de numérisation [JOHNSTONE ET AL., 2004]

Ces méthodes traitent habituellement chaque forme d'onde de façon **indépendante des autres**.

Je vais expliquer seulement une d'entre elles qui est souvent utilisée. Il s'agit de la méthode par **détection de pics**. Le but de cette méthode est de déterminer la bathymétrie en détectant les pics de surface et du fond de l'eau de la forme d'onde LiDAR. On peut ainsi estimer la bathymétrie en multipliant la différence de temps entre les deux pics par la vitesse de la lumière dans l'eau.

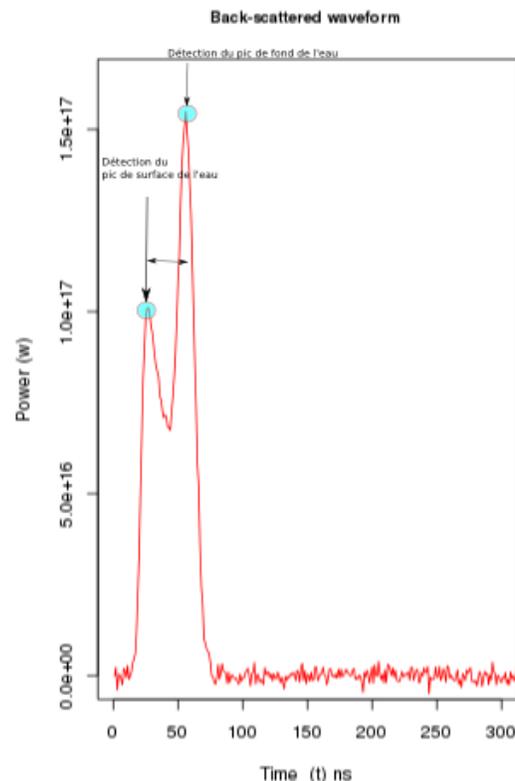


FIGURE 7.1 – Méthode d'inversion classique par détection de pics

1.2 Les limites des méthodes d'inversion de formes d'onde actuelles

Les méthodes proposées dans la littérature sont inefficaces dans certain cas. Par exemple, lorsque nous appliquons ces méthodes :

- dans le **cas des très faibles profondeurs** : la difficulté réside dans le fait que nous ne pouvons plus distinguer le pic de surface de celui du fond de l'eau car les formes d'onde présentent un seul pic qui contient le pic de surface et celui du fond.

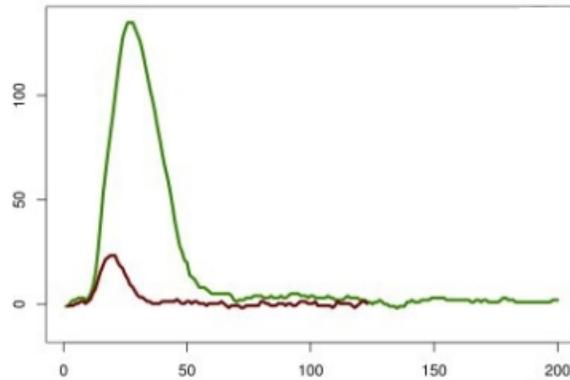


FIGURE 7.2 – On ne distingue plus la composante de surface (courbe rouge) et celle du fond de l'eau (courbe verte) (Référence [76])

- dans le cas des **eaux turbides** où l'on observe difficilement le pic de fond du l'eau.¹

1. Généralement à cause des phénomènes d'absorption dans l'eau

2 Vers de nouvelles méthodes d'inversion de formes d'onde LiDAR

Les méthodes d'inversion que j'ai utilisées proposent une approche totalement différente par rapport aux méthodes précédentes.

Le but est de construire des modèles statistiques². Le but étant d'estimer la bathymétrie³ (variable environnementale qui nous intéresse) à partir d'un échantillon de formes d'onde. La figure 7.3 illustre la méthode que nous allons utiliser pour estimer la bathymétrie de manière statistique.

Les étapes de nos méthodes (Figure 7.3) sont les suivantes :

Étape 1 : Je choisis une base de données de N formes d'onde LiDAR. Chaque forme d'onde est caractérisée par une bathymétrie (ou une hauteur d'arbre) mesurée sur le terrain. On peut la noter Z_{mes}^i (i le numéro de la forme d'onde).

Étape 2 : On détermine des variables explicatives β_j^i (l'indice j correspond au numéro de la variable explicative et i au numéro de la forme d'onde) à l'aide de quatre méthodes différentes (l'Analyse en Composantes Principales, les ondelettes de Haar, les splines cubiques et les B-Splines).

Étape 3 : On sélectionne aléatoirement K=80% de nos N formes d'onde. Grâce à ces K formes d'onde, on peut écrire l'équation suivante :

$$\begin{bmatrix} Z_{mes}^1 \\ Z_{mes}^2 \\ Z_{mes}^3 \\ \vdots \\ \vdots \\ Z_{mes}^{K-1} \\ Z_{mes}^K \end{bmatrix} = x_1 \begin{bmatrix} \beta_1^1 \\ \beta_1^2 \\ \beta_1^3 \\ \vdots \\ \vdots \\ \beta_1^{K-1} \\ \beta_1^K \end{bmatrix} + \dots + x_j \begin{bmatrix} \beta_j^1 \\ \beta_j^2 \\ \beta_j^3 \\ \vdots \\ \vdots \\ \beta_j^{K-1} \\ \beta_j^K \end{bmatrix} + \dots + x_p \begin{bmatrix} \beta_p^1 \\ \beta_p^2 \\ \beta_p^3 \\ \vdots \\ \vdots \\ \beta_p^{K-1} \\ \beta_p^K \end{bmatrix} + \epsilon \quad (7.1)$$

ϵ est une constante, i est le numéro de la forme d'onde, j est le numéro de la variable explicative, p est le nombre des variables explicatives.

Étape 4 : On détermine ensuite les coefficients (x_1, x_2, \dots, x_p) de l'équation 7.1 pour construire notre modèle de prédiction à partir des K formes d'onde. On construit par exemple le modèle prédictif suivant :

$$Z_{est}^i = \sum_{j=1}^p x_j \beta_j^i + \epsilon \quad (7.2)$$

avec i le numéro de la forme d'onde, j le numéro de la variable explicative, p le nombre de variables explicatives et ϵ une constante.

Étape 5 : Pour valider le modèle, on applique ce modèle prédictif sur les 20% des formes d'onde restantes. On peut également utiliser les arbres de décision (CART) et les arbres aléatoires (Random Forest).

Étape 6 : On peut ainsi comparer la bathymétrie mesurée sur le terrain et la bathymétrie que nous avons estimée avec le modèle construit.

2. En utilisant la régression linéaire multiple, les arbres décisionnels (CART) et les arbres aléatoires (Random Forest)

3. Ou la hauteur des arbres

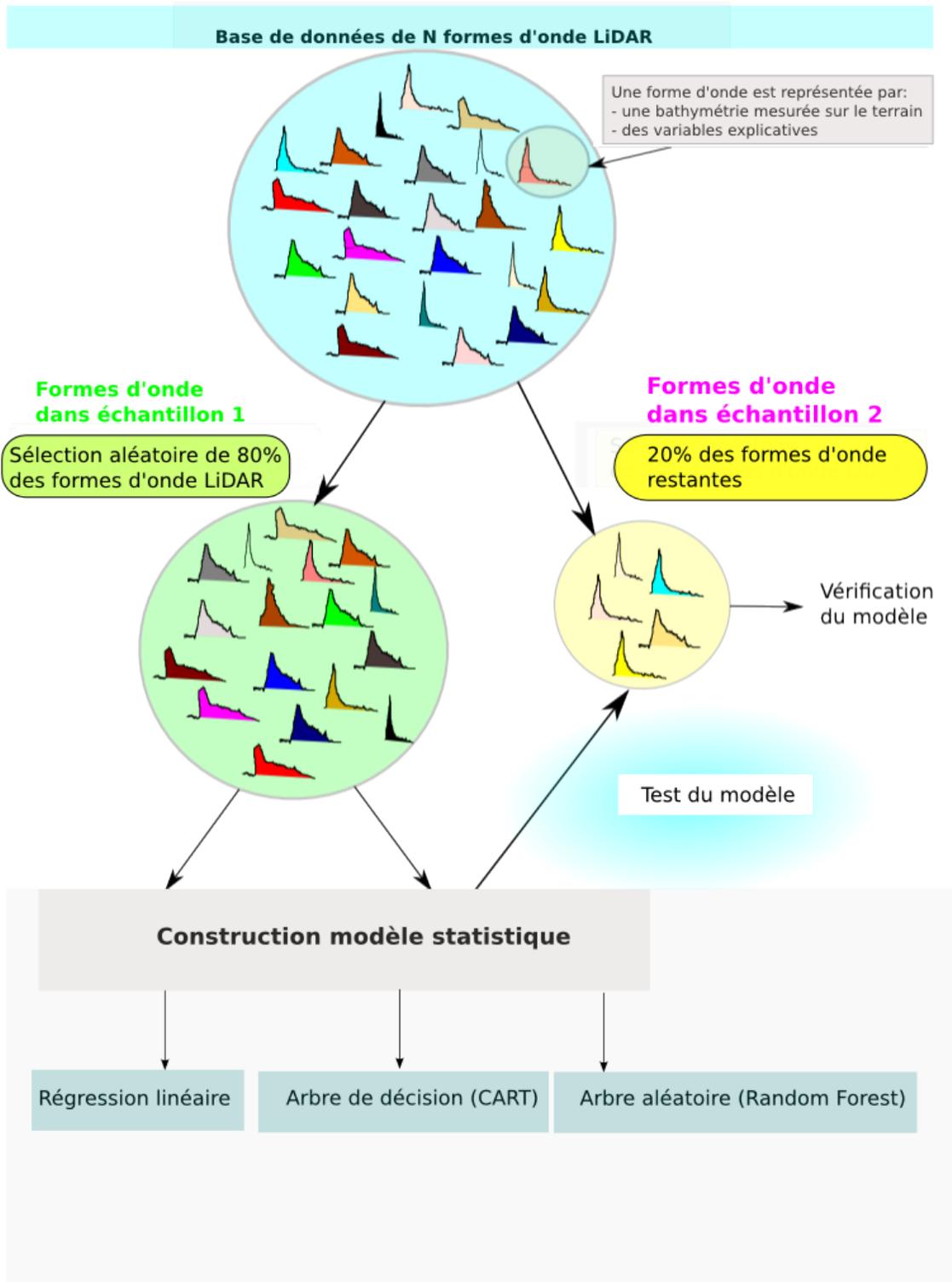


FIGURE 7.3 – Méthodologie pour les nouvelles méthodes d'inversion de formes d'onde LiDAR. Estimation de la bathymétrie et de la hauteur des arbres en utilisant une approche statistique.

Chapitre 8

Inversion des formes d'onde LiDAR avec l'Analyse en Composantes Principales

Sommaire

| | | |
|----------|---|-----------|
| 1 | Rappel sur l'Analyse en Composantes Principales | 62 |
| 1.1 | Principe de l'Analyse en Composantes Principales | 62 |
| 1.2 | Type de données pour effectuer une Analyse en Composantes Principales | 62 |
| 1.3 | Les Composantes Principales | 63 |
| 1.4 | Résumé sur l'Analyse en Composantes Principales | 64 |
| 2 | Inversion des formes d'onde LiDAR à l'aide de l'Analyse en Composantes Principales | 65 |

Résumé du chapitre : Dans ce chapitre, j'explique en première partie les notions de base de l'Analyse en Composantes Principales¹ (ACP). J'effectue ensuite une ACP sur la base de données du Golfe du Morbihan² avec le langage **R** en utilisant le package **FactoMineR**. La fonction **PCA()** du package **FactoMineR** me permet ainsi de déterminer les composantes principales³. Ces composantes principales représentent les variables explicatives des formes d'onde (Figure 7.3) et me permettent de construire des modèles statistiques (Régression linéaire multiple, arbres de décisions et les arbres aléatoires). Dans ce document, j'explique la démarche pour la construction d'un modèle de régression linéaire le plus simple. J'expose ensuite directement les résultats dans un tableau pour les différents modèles statistiques construits sur la base de données du Golfe du Morbihan. Je commente les résultats des modèles d'inversion des formes d'onde LiDAR de toutes les bases de données dans le chapitre 12.

1. Sans entrer dans les détails mathématiques
 2. acquis avec le capteur APD (Avalanche PhotoDiode)
 3. cf annexe L

1 Rappel sur l'Analyse en Composantes Principales

1.1 Principe de l'Analyse en Composantes Principales

L'Analyse en composantes Principales (ACP) est une méthode de **statistique exploratoire et prédictive** permettant de décrire un grand tableau de données de type individu \times variables. Lorsque les individus sont décrits par un nombre important de variables, aucune représentation graphique simple ne permet de visualiser le nuage de points formé par les données. L'ACP propose une représentation dans un **espace de dimension réduite**, permettant ainsi de mettre en évidence d'éventuelles structures au sein des données. Pour cela, l'ACP transforme des variables liées entre elles (ou corrélées) en nouvelles variables indépendantes les unes des autres (non corrélées ou "orthogonales" entre elles). Ces variables sont appelées "**composantes principales**" et sont classées par ordre "d'importance" : la première composante principale est l'axe qui capte le plus de variabilité du tableau. Il s'agit d'une approche à la fois **géométrique** (représentation des variables dans un nouvel espace selon des directions d'inertie maximale) et **statistique** (recherche d'axes indépendants expliquant au mieux la variabilité, la variance⁴ des données).

1.2 Type de données pour effectuer une Analyse en Composantes Principales

L'ACP s'applique à des **données quantitatives** représentées sous la forme d'un **tableau de données à N lignes et p colonnes** (Figure 8.1). Chaque ligne correspond à une forme d'onde pour $i=1, \dots, N$ et chaque forme d'onde (individu) est constituée de p samples (ou points de mesures).

| Variables | | X^1 | ... | X^j | ... | X^p |
|------------------------------|-------|-------|-----|---------|-----|-------|
| Individus (Formes d'onde) | X_1 | | | ⋮ | | |
| | ⋮ | | | ⋮ | | |
| | X_i | ... | ... | X_i^j | ... | ... |
| | ⋮ | | | ⋮ | | |
| | X_N | | | ⋮ | | |

FIGURE 8.1 – Représentation des données

On considère sur la figure (8.1) : N formes d'onde X_1, \dots, X_N sur lesquels on a p samples $X^1, \dots, X^j, \dots, X^p$ (j désignant le numéro du sample). Chaque ligne i du tableau correspond à une forme d'onde X_i constituée de p points de mesures de la base de données. Par exemple, pour la base de données du Golfe du Morbihan⁵, on a $N = 189$ formes d'onde et chaque forme d'onde est constituée de $p = 64$ samples (ou points de mesures).

4. Définition : En statistique, la **variance** est une mesure servant à caractériser la dispersion d'un nuage de points. Elle indique de quelle manière le nuage se disperse autour de sa moyenne. Une variance de zéro signale que toutes les valeurs sont identiques. Une petite variance est signe que les valeurs sont proches les unes des autres alors qu'une variance élevée est signe que celles-ci sont très écartées (Wikipédia).

5. avec le capteur Avalanche PhotoDiode

1.3 Les Composantes Principales

L'Analyse en Composantes Principales définit de nouvelles variables (changement de base) :

$$PC_i^j = \alpha_1 X_i^1 + \dots + \alpha_j X_i^j + \dots + \alpha_p X_i^p \quad i = 1, \dots, N \quad j = 1, \dots, p \quad (8.1)$$

telles que :

- PC_i^1 est la variable la plus corrélée aux variables X_i^j de la forme d'onde n°i.
- de même pour PC_i^2 , avec PC_i^2 orthogonale à PC_i^1
- et ainsi de suite.

avec PC_i^j la nouvelle variable n°j de la forme d'onde n°i. Pour plus d'informations sur la construction des composantes principales cf. Annexe [L](#)

1.4 Résumé sur l'Analyse en Composantes Principales

L'Analyse en Composantes Principales nous permet de :

- représenter les individus en fonction de leur "proximité" (ou ressemblance)

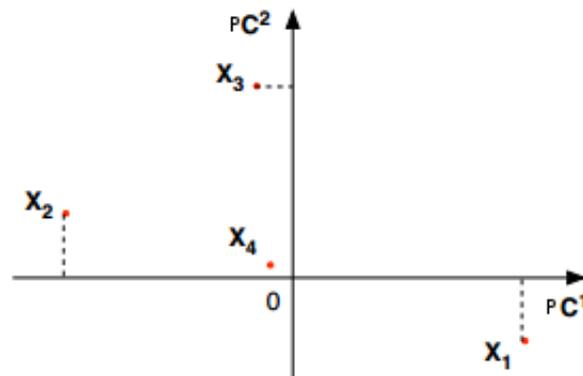


FIGURE 8.2 – Graphique des individus

- en transformant les p anciennes variables par p nouvelles variables :

| Variables | X^1 | ... | X^j | ... | X^p |
|------------------------------|-------|-----|-------|---------|-------|
| Individus (Formes d'onde) | X_1 | | | | |
| | ⋮ | | | | |
| | X_i | ... | ... | X_i^j | ... |
| | ⋮ | | | | |
| | X_N | | | | |

→

| Nouvelles variables | PC^1 | ... | PC^j | ... | PC^p |
|---------------------|--------|-----|--------|---------|--------|
| Individus | X_1 | | | | |
| | ⋮ | | | | |
| | X_i | ... | ... | U_i^j | ... |
| | ⋮ | | | | |
| | X_N | | | | |

FIGURE 8.3 – Tableau des données avec les p variables initiales .

FIGURE 8.4 – Tableau des données avec les p nouvelles variables. Changement de base.

Nous ferons le choix de garder **les premières composantes principales** qui conservent **la grande majorité de l'information** contenue dans le tableau de données initiales (Figure 8.3).

2 Inversion des formes d'onde LiDAR à l'aide de l'Analyse en Composantes Principales

Dans cette partie, nous utiliserons dans une première étape l'Analyse en Composantes Principales pour au préalable explorer nos données. Nous essaierons de voir s'il existe une corrélation entre la bathymétrie mesurée sur le terrain et les nouvelles composantes principales⁶. Effectuons une Analyse en Composantes Principales sur le jeu de données de N formes d'onde LiDAR bathymétriques (réelles) correspondant au site du Golfe du Morbihan⁷ afin de construire un modèle de régression linéaire simple à partir d'un sous échantillon de $K = 80\%$ de N formes d'onde qui permettra d'estimer la bathymétrie et de le tester sur un sous échantillon de $N - K$ formes d'onde⁸ (Figure 7.3). Ce processus est répété 5 fois⁹.

Plusieurs fonctions de **R** permettent de faire une **ACP**. Nous conseillons la fonction **PCA** du package **FactoMineR**.

Nous rappelons la structure du tableau sur lequel nous allons effectuer l'ACP (Tableau de données 8.1).

| - | Bathymétrie mesurée | $P(t_1)$ | $P(t_2)$ | ... | ... | $P(t_{64})$ |
|--------------------|---------------------|----------|----------|-----|-----|-------------|
| Forme d'onde n°1 | 0.9288 | 8 | 7 | .. | ... | 7 |
| Forme d'onde n°2 | 0.6158 | 8 | 9 | ... | ... | 6 |
| ... | ... | ... | ... | ... | ... | ... |
| Forme d'onde n°188 | 2.1400 | 9 | 9 | ... | ... | 5 |
| Forme d'onde n°189 | 2.3797 | 7 | 7 | ... | ... | 9 |

TABLE 8.1 – Tableau de 189 formes d'onde enregistrées par un lidar bathymétrique avec le capteur APD. Les lignes correspondent aux individus (Formes d'onde) et les colonnes aux variables (Puissance rétrodiffusée détectée par le LiDAR à un instant donné t_j ($j=1,\dots,64$)). La colonne "bathymétrie mesurée" correspond à une **variable supplémentaire**. Elle n'interviendra pas dans la construction des composantes principales. Elle nous permettra de voir la corrélation avec les composantes principales et de construire des modèles statistiques permettant d'estimer la bathymétrie.

Les étapes de l'ACP avec la package FactoMineR

A- Importer le jeu de données

La première étape consiste à importer notre jeu de données dans **R**.

```
1 tableau_jeu_donnees=read.csv(file="toutGAPDechantillonne.csv",sep=" ",dec=".",header=F)
```

B- Choix des variables actives

Le choix des variables actives est important : Ce sont ces variables et uniquement ces variables qui vont participer à la construction des axes de l'ACP. Les variables actives sont les colonnes $P(t_j)$ du tableau 8.1 pour $j=1,\dots,64$.

C- Standardiser ou non les variables

Lors d'une ACP nous pouvons centrer-réduire les variables ou **seulement les centrer**. Dans notre cas, nous devons centrer les données et ne surtout pas réduire les variables car les unités sont identiques (Cette étape est cruciale).

6. variables explicatives

7. acquis avec le capteur APD Avalanche PhotoDiode

8. sur les 20% des formes d'onde restante

9. 5-folds

Pour réaliser l'ACP à l'aide du package **FactoMineR**, il faut charger le package préalablement installé. Nous utilisons ensuite la fonction **PCA()**.

```
1 acp=PCA(tableau_jeu_donnees[,c(2:65)], scale.unit=FALSE, ncp=length(2:65), graph=T, quali.sup=1)
```

Pour éviter, la réduction des variables avec la fonction **PCA()** on utilise l'argument **scale.unit=FALSE**. La variable 1 (la bathymétrie mesurée) est une variable supplémentaire (**quali.sup=1**).

D- Choisir le nombre d'axes (ou de composantes principales)

Plusieurs solutions existent pour déterminer le nombre d'axes à retenir en ACP. Pour cela, nous regardons la quantité d'information recueillie par chaque composantes principales (Figure 8.5 et le tableau 8.2) (ou axes factoriels). On peut représenter le diagramme des inerties associées à chaque axe avec la commande :

```
1 barplot(acp$eig[,2], names=paste("PC", 1:nrow(acp$eig)))
```

Pourcentage de variance expliquée par les axes factoriels (%)

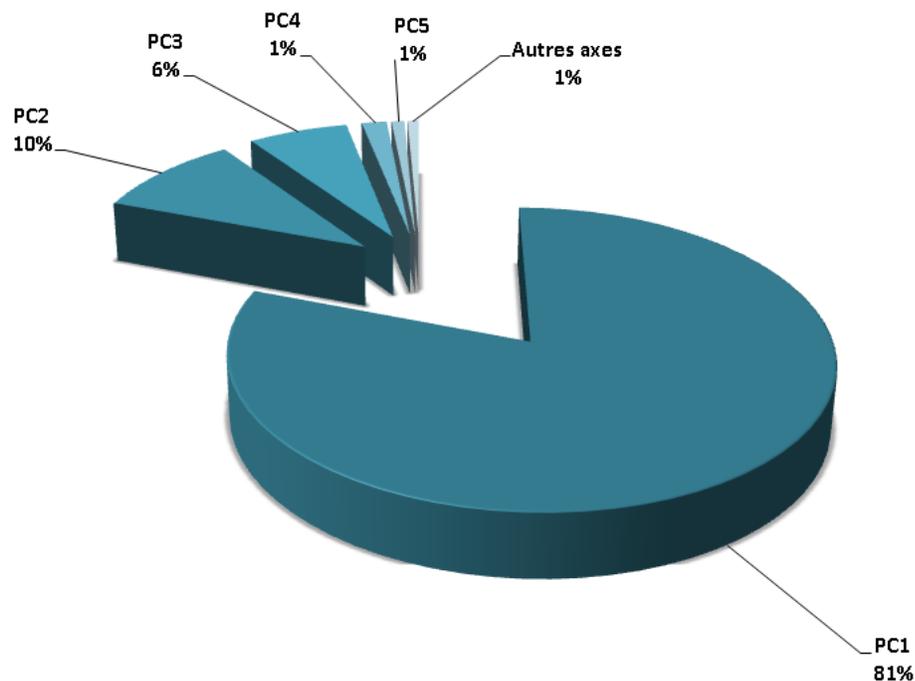


FIGURE 8.5 – Quantité d'information recueillie par les composantes principales (ou axes factoriels)

CHAPITRE 8. INVERSION DES FORMES D'ONDE LIDAR AVEC L'ANALYSE EN COMPOSANTES PRINCIPALES

Le tableau ci-dessous, regroupe les valeurs propres, le pourcentage d'inertie et le pourcentage cumulé expliquée par les 5 premiers axes :

| Composantes principales | Valeurs propres | Pourcentage de variance (%) | Pourcentage de variance cumulé(%) |
|-------------------------|-----------------|-----------------------------|-----------------------------------|
| PC 1 | 21566.23 | 80,74 | 80.74 |
| PC 2 | 2442.48 | 9,73 | 90.47 |
| PC 3 | 2119.89 | 6,4 | 96.88 |
| PC 4 | 1009.20 | 1,6 | 98.48 |
| PC 5 | 233.07 | 0,83 | 99.20 |
| Autres axes | 233.07 | 0.71 | 100 |

TABLE 8.2 – Tableau des valeurs propres, du pourcentage de variance et de la variance cumulée des axes principaux ("PC : Principal Component").

Le **premier axe (PC1)** expriment **80.74 %** (Figure 8.5 et tableau 8.2) de l'inertie totale, c'est à dire que **80.74 %** de l'information du tableau de données est contenue dans la première composante principale. Ici, les cinq premières composantes permettent d'expliquer **99.20 %** de l'inertie totale.

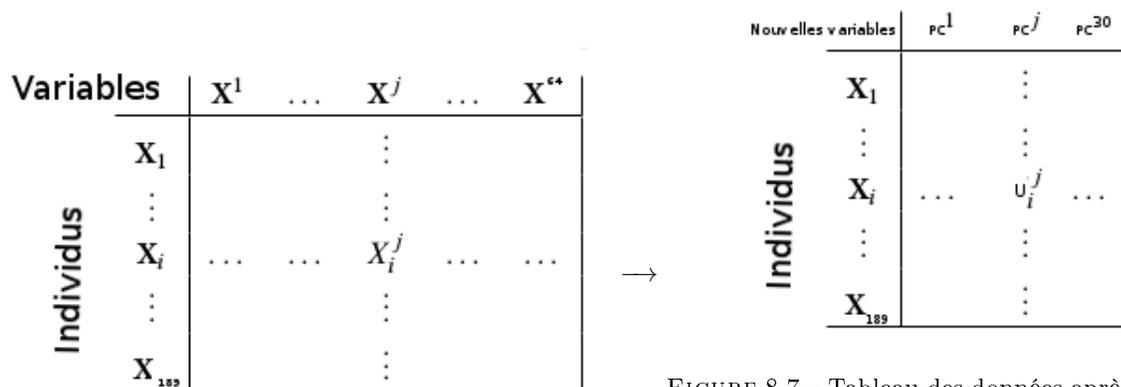


FIGURE 8.6 – Tableau des données avec $p=64$ variables initiales et 189 formes d'onde.

FIGURE 8.7 – Tableau des données après changement de base. Nous avons fait le choix de conserver seulement les 30 premières nouvelles variables. **Les 5 premières composantes principales résument 99.20 % de l'information initiale du tableau.**

Sur les $p=64$ nouvelles variables nous avons conservé les 30 premières car elles permettent à elles seules d'expliquer une part très importante de l'information contenue dans le tableau initial (plus de 99.20% soit **moins de 1% de perte d'information**). Nous avons donc réduit la dimension du tableau initial en faisant le choix de conserver les 30 premières nouvelles variables. Une forme d'onde peut donc être décrite par 30 nouvelles variables au lieu des 64 variables initiales avec une perte d'information de moins de 1%. Nous avons ainsi **réduit la dimension de notre problème**.

E- Analyse des résultats avec le graphique des individus

La fonction **PCA** fournit par défaut le **graphique des individus** pour les deux premières composantes (Figure 8.8). Chaque point de la figure 8.8 correspond à une forme d'onde réelle. Chaque forme d'onde est représentée par un numéro (de 1 à 189). Chaque forme d'onde est décrite par la première et deuxième composantes principales.

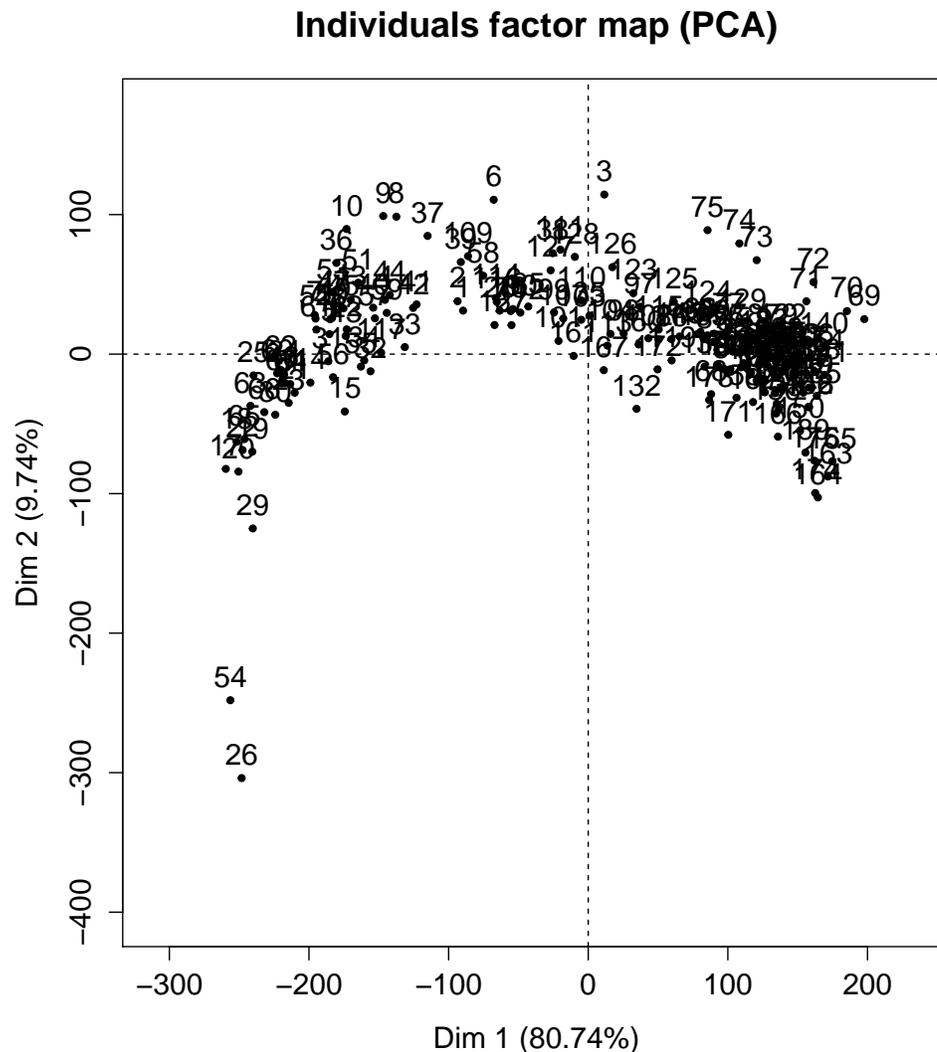


FIGURE 8.8 – Graphique des individus Dim 1 et 2 représentent respectivement la première et deuxième composantes principales.

Sur la figure 8.8, les formes d'onde qui se "ressemblent" (ou qui sont "similaires") du point de vue de la première et de la deuxième composante principale sont proches entre elles. Et réciproquement, les formes d'onde qui se "différencient" du point de vue de la première et de la deuxième composante principale seront éloignées entre elles.

Essayons de voir s'il existe un lien entre la bathymétrie mesurée sur le terrain (variable supplémentaire) et les composantes principales. Pour cela, nous allons ré-afficher la figure 8.8 :

- en coloriant en noir les formes d'onde qui ont une bathymétrie entre [0 m,1.5 m] (Figure 8.9)
- en coloriant en rouge les formes d'onde qui ont une bathymétrie comprise entre [1.5m ,3 m] (Figure 8.9)

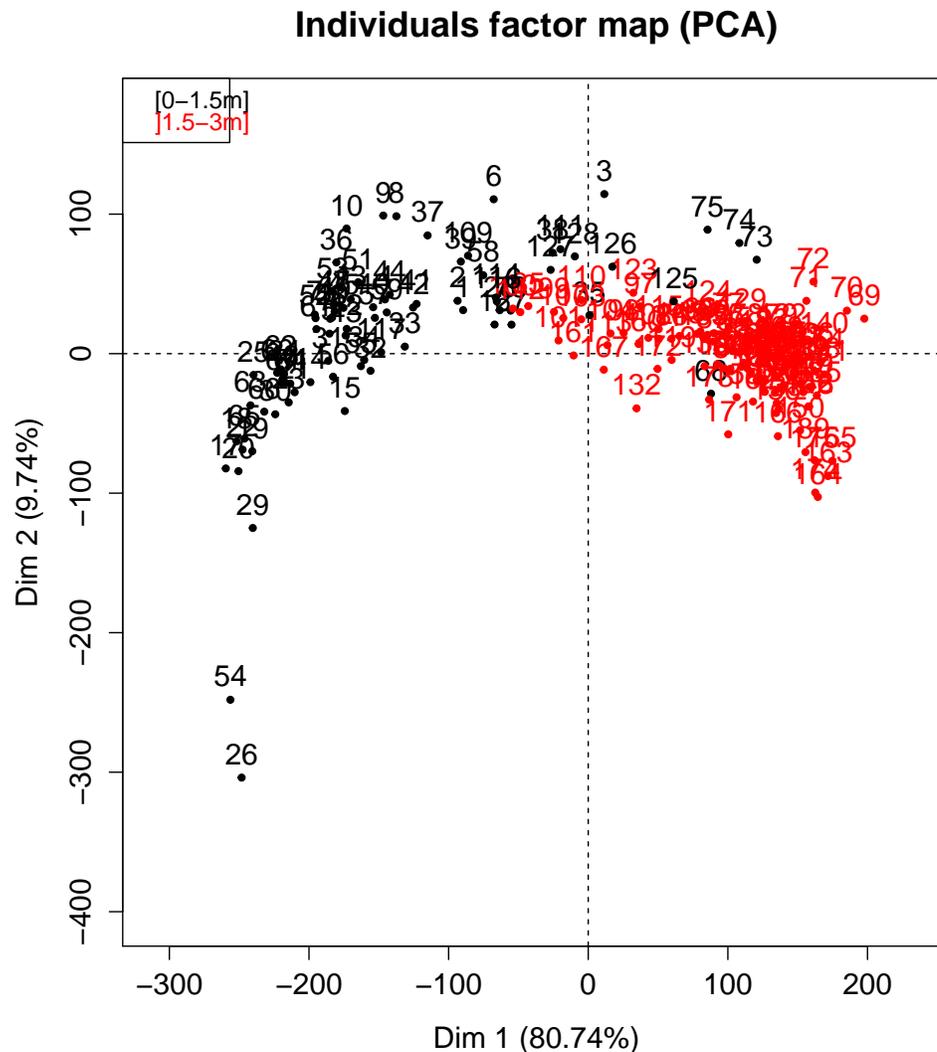


FIGURE 8.9 – Graphique des individus (ou des formes d'onde) en fonction de deux gammes de profondeur.

La figure 8.9 illustre deux groupes de formes d'onde, un premier à gauche (en noir) et un second à droite (en rouge).

F- Corrélation entre la bathymétrie mesurée sur le terrain et la première composante principale

Essayons de voir s'il existe une corrélation entre la **première composante principale (PC1)** et la **bathymétrie**. Pour cela, on représente la bathymétrie en fonction de la première composante principale (PC1).

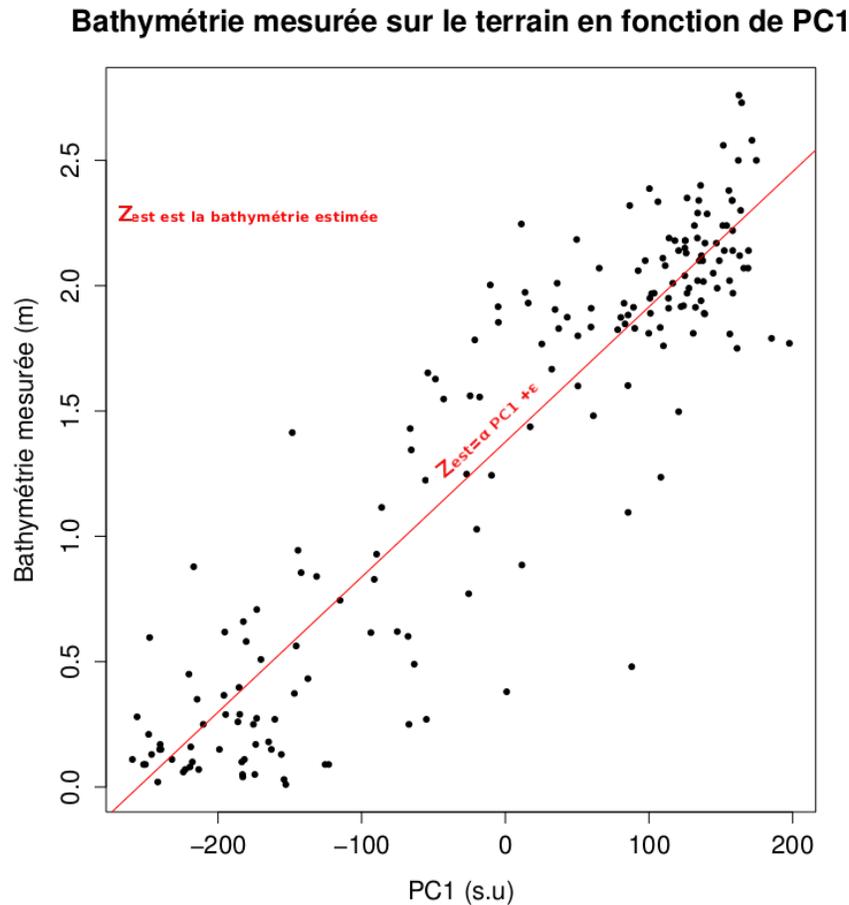


FIGURE 8.10 – Régression linéaire de la forme : $Z_{est} = \alpha PC1 + \epsilon$ avec $\alpha = 0.005391$ et $\epsilon = 1.376486$

Sur la figure 8.10, on peut s'apercevoir que le nuage de points est réparti uniformément dans une ellipse. On note donc une corrélation entre PC1 et la bathymétrie mesurée sur le terrain. On peut donc construire un modèle de régression linéaire de la forme (Figure 8.10) :

$$Z_{est} = \alpha PC1 + \epsilon \tag{8.2}$$

avec

- Z_{est} est la "**variable d'intérêt**", Il s'agit de la bathymétrie estimée.
- α et ϵ sont les coefficients du modèle
- PC1 est la "**variable explicative**" correspondant aux coordonnées des individus (formes d'onde) dans le nouvel espace selon la première composante principale.

Le coefficient de détermination du modèle de régression de la figure 8.10 est de $R^2 = 84.6\%$ avec un Root Mean Squarre Error ¹⁰ (RMSE) de 32.78 cm.

| Modèle statistique | R^2 (%) | RMSE (cm) |
|-----------------------------------|-----------|-----------|
| $Z_{est} = \alpha PC1 + \epsilon$ | 84.7% | 33 cm |

Nous avons créé un modèle de régression linéaire simple à partir de N (=189) formes d'onde. Nous avons ensuite employé plusieurs modèles de régression en utilisant la validation croisée (cross-validation k-folds). Nous regroupons le coefficient de détermination R^2 et le RMSE dans le tableau selon les différents modèles statistiques et méthodes utilisées ¹¹ :

10. Erreur quadratique moyenne

11. pour la base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode

| Modele statistique | Méthode utilisée | R^2 (%) | RMSE (m) | Nbre de variables descriptifs |
|---|--------------------|-----------|----------|-------------------------------|
| $Z_{est} = \alpha PC1 + \epsilon$ | RLM | 84,7 | 0,33 | 1 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \epsilon$ | RLM | 85.65 | 0.32 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \epsilon$ | RLM | 85.85 | 0.31 | 3 |
| $Z_{est} = \alpha PC1 + \beta W_{ext} + \epsilon$ | RLM | 84.97 | 0.32 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \beta W_{ext} + \epsilon$ | RLM | 85.98 | 0.31 | 3 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \beta W_{ext} + \epsilon$ | RLM | 85.44 | 0.32 | 4 |
| $Z_{est} = \alpha PC1 + \epsilon$ | CART | 80.94 | 0.37 | 1 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \epsilon$ | CART | 83.88 | 0.34 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \epsilon$ | CART ¹² | 84.58 | 0.33 | 3 |
| $Z_{est} = \alpha PC1 + \beta W_{ext} + \epsilon$ | CART | 81.81 | 0.36 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \beta W_{ext} + \epsilon$ | CART | 85.59 | 0.32 | 3 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \beta W_{ext} + \epsilon$ | CART | 83.37 | 0.34 | 4 |
| $Z_{est} = \alpha PC1 + \epsilon$ | RF | 79.83 | 0.38 | 1 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \epsilon$ | RF | 88.15 | 0.29 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \epsilon$ | RF | 85.54 | 0.32 | 3 |
| $Z_{est} = \alpha PC1 + \beta W_{ext} + \epsilon$ | RF | 80.43 | 0.37 | 2 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \beta W_{ext} + \epsilon$ | RF | 84.44 | 0.33 | 3 |
| $Z_{est} = \alpha_1 PC1 + \alpha_2 PC2 + \alpha_3 PC3 + \beta W_{ext} + \epsilon$ | RF | 85.71 | 0.32 | 4 |
| $Z_{est} = \sum \alpha_n PCn + \epsilon$ | Stepwise+RLM | 87.38 | 0.3 | 13 |
| $Z_{est} = \sum \alpha_n PCn + \epsilon$ | Stepwise+CART | 83.16 | 0.34 | 13 |
| $Z_{est} = \sum \alpha_n PCn + \epsilon$ | Stepwise+RF | 83.16 | 0.34 | 13 |
| $Z_{est} = \sum \alpha_n PCn + \beta W_{ext} + \epsilon$ | Stepwise+RLM | 87.38 | 0.3 | 14 |
| $Z_{est} = \sum \alpha_n PCn + \beta W_{ext} + \epsilon$ | Stepwise+CART | 83.16 | 0.34 | 14 |
| $Z_{est} = \sum \alpha_n PCn + \beta W_{ext} + \epsilon$ | Stepwise+RF | 83.16 | 0.34 | 14 |

TABLE 8.3 – Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables descriptifs des différents modèles mathématiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : Random Forest, **stepwise** : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 30 premières variables, W_{ext} est la variable qui correspond à la partie utile d'une forme d'onde LiDAR.)

Chapitre 9

Inversion des formes d'onde LiDAR avec l'analyse en ondelettes

Sommaire

| | | |
|----------|--|-----------|
| 1 | Rappel : Ondelettes de Haar | 74 |
| 1.1 | Principe de l'analyse en ondelettes | 74 |
| 1.2 | La base de Haar | 74 |
| 1.3 | Les coefficients d'ondelettes | 76 |
| 1.4 | Reconstruction d'un signal de forme d'onde à l'aide des ondelettes de Haar | 76 |
| 2 | Inversion des formes d'onde LiDAR avec l'analyse en ondelettes | 78 |

Résumé du chapitre : Dans ce chapitre, je fais un rappel de l'analyse en ondelettes. Cette méthode nous permet d'extraire **les coefficients d'ondelettes** représentant les variables explicatives pour la construction des modèles statistiques en utilisant la même approche que j'ai utilisée avec l'Analyse en Composantes Principales. J'explique ensuite toujours sur la base de données du Golfe du Morbihan¹ comment extraire les coefficients d'ondelettes sur une forme d'onde pour pouvoir dans une seconde étape faire une inversion des formes d'onde en utilisant la fonction `stepwise()` de **R** et construire mes modèles statistiques. Je commente les résultats des modèles d'inversion des formes d'onde LiDAR de toutes les bases de données dans le chapitre 12.

1. avec le capteur Avalanche PhotoDiode

1 Rappel : Ondelettes de Haar

1.1 Principe de l'analyse en ondelettes

Les ondelettes peuvent être utilisées afin de **décomposer une fonction réelle (signal) périodique ou non périodique en une combinaison linéaire de fonctions élémentaires avec des niveaux de résolution de plus en plus fins** (Figure 9.1). Le principe de cette théorie repose sur le travail de Fourier relatif à la décomposition d'un signal sur un ensemble de fonctions sinusoïdales. Il existe plusieurs familles d'ondelettes (Haar, Daubechies...). Dans mon cas, j'utilise **les ondelettes de Haar** pour extraire **les variables explicatives**, c'est à dire **les coefficients d'ondelettes** pour construire mes modèles statistiques. Si on fait le parallèle avec l'Analyse en Composante Principale, au lieu d'effectuer une régression linéaire sur les composantes principales, je fais cette fois-ci une régression linéaire sur les coefficients d'ondelettes que l'on calcule à l'aide de la fonction `dwt()` du package `waveslim`.

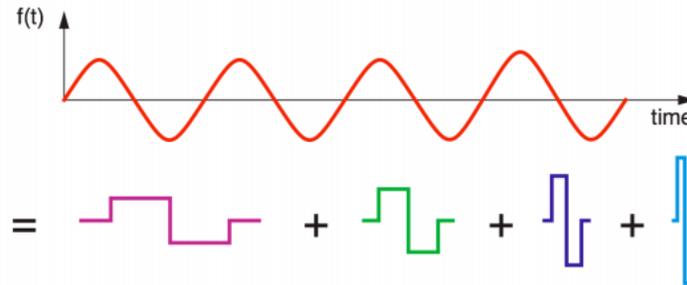


FIGURE 9.1 – Ondelettes de Haar - Fonction $f(t)$ décomposée en un ensemble de fonctions élémentaires discontinues avec des niveaux de résolution de plus en plus fins.

1.2 La base de Haar

L'idée de l'analyse en ondelettes est de décomposer un **signal de $N = 2^k$ points**² sur une base de fonctions discontinues et orthogonales entre elles issue des translations et dilatations de fonctions ϕ et ψ appelées respectivement **fonction d'échelle père** et **ondelette mère**. La famille $\{\phi, \psi\}$ est appelée base de Haar.

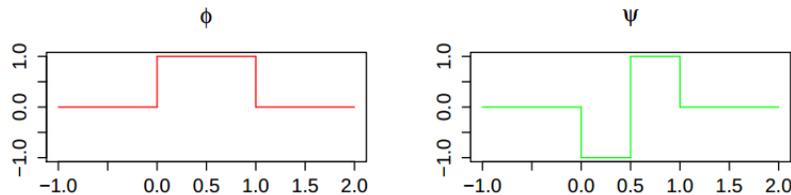


FIGURE 9.2 – Fonction d'échelle père ϕ et ondelette mère ψ

Prenons une forme d'onde LiDAR réelle du Golfe du Morbihan, c'est à dire un signal de $N=64$ samples (ou points de mesures) et calculons le nombre de niveaux de résolution k . Nous avons $N = 64 \text{ points} = 2^k \Rightarrow k = 6$.

Un signal à $N = 2^6 = 64$ samples peut-être décomposé sur 6 niveaux et chaque niveau regroupe des fonctions de base avec des niveaux de résolution différents.

2. Pour utiliser les ondelettes, il faut obligatoirement avoir un signal de 2^k points.

- Le **niveau 1** est composé de **2 fonctions d'échelles pères** de même taille (qui permettent d'approximer le signal le plus grossièrement).
- Le **niveau 2** est composé de **2 fonctions d'ondelettes mères** de même taille.
- Le **niveau 3** est composé de **4 fonctions d'ondelettes mères** de même taille 2 fois plus comprimées que les fonctions d'ondelettes mère du niveau 2.
- Le **niveau 4** est composé de **8 fonctions d'ondelettes mères** de même taille 2 fois plus comprimées que les fonctions d'ondelettes mère du niveau 3.
- Le **niveau 5** est composé de **16 fonctions d'ondelettes mères** de même taille 2 fois plus comprimées que les fonctions d'ondelettes mère du niveau 4.
- Le **niveau 6** est composé de **32 fonctions d'ondelettes mères** de même taille 2 fois plus comprimées que les fonctions d'ondelettes mère du niveau 5.

Notons aussi que nous possédons **autant de fonctions de base que de points** qui constitue le signal. Ici, pour un signal avec **64 points nous avons 64 fonctions de base** (et 64 coefficients d'ondelettes³). La figure 9.3 illustre toutes les fonctions de base nécessaires à **la reconstruction complète d'un signal** composé de 64 points et regroupées sur plusieurs niveaux de résolution. La figure indique notamment le nombre de fonctions de base qui diffère suivant les niveaux.

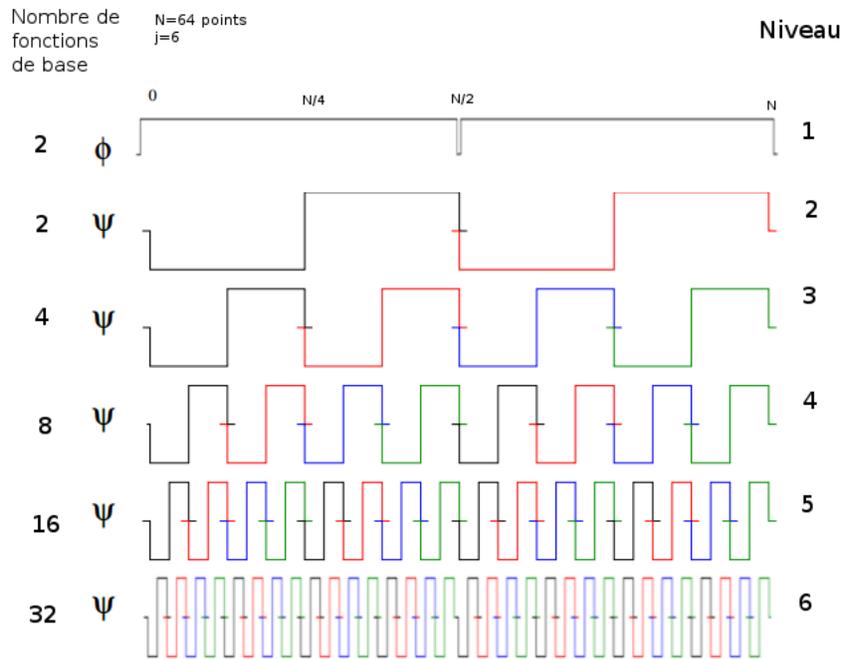


FIGURE 9.3 – Les fonctions de base avec des niveaux de résolution de plus en plus fin. Les fonctions de base des deux premiers niveaux ont toujours chacune 2 fonctions de base. Le premier niveau correspond toujours aux **fonctions d'échelle père**, elles approximent le signal grossièrement. Les autres niveaux sont des **fonctions d'ondelette mère** qui sont 2 fois plus nombreuses et plus petites lorsque le niveau augmente. On répète le processus jusqu'à avoir 64 fonctions de base au total.

3. Ce sont les coefficients de la combinaison linéaire des fonctions de base.

1.3 Les coefficients d'ondelettes

Notons qu'un signal échantillonné avec 64 points comporte 64 coefficients d'ondelettes (et 64 fonctions de base). Les fonctions de bases sont représentées sur la figure 9.3. On peut calculer les coefficients d'ondelettes correspondant à chaque fonction de base de la figure 9.3 à l'aide de **l'algorithme de Mallat**⁴. Dans **R**, nous pouvons faire appel à la fonction `dwt` (Discret Wavelet Transform) du package `waveslim` qui utilise l'algorithme de Mallat pour calculer les coefficients d'ondelettes .

```

1 a=dwt(Signal, 'haar', (log2(length(Signal))/log2(2) )-1 )
2
3 # Signal : Forme d'onde discrete de 64 points
4 # Choix de la famille d'ondelettes : "Haar"
5 # (log2(length(Signal))/log2(2) )-1 : Indice k=ln(N)/ln(2)

```

La figure 9.4 illustre les différents niveaux des coefficients d'ondelettes pour une forme d'onde correspondant **dans l'ordre** à chaque fonction de base de la figure 9.3. Les niveaux 1,2,3,4,5 et 6 de la figure 9.4 sont respectivement appelés par la fonction `dwt()` niveau : **s5**, **d5**, **d4**, **d3**, **d2** et **d1** (s pour les coefficients de la fonction d'échelle (scale) et **d** pour les coefficients en ondelettes mères, ce que l'on nomme "les détails").

```

> dwt(Z, 'haar', (log2(length(Z))/log2(2) )-1)

$s5
[1] 418.25366 68.58936

$d5
[1] 138.94648 -22.62742

$d4
[1] 159.5 -46.5 -23.0 -3.0

$d3
[1] 2.1213203 89.0954544 -10.2530483 -47.0226009 -21.5667568 -1.0606602 -0.3535534 -1.0606602

$d2
[1] 1.5 1.5 56.5 5.5 -7.5 2.0 -10.0 -21.5 -14.0 -2.5 -1.5 -1.0 1.0 1.5 -2.5 -2.0

$d1
[1] 1.4142136 0.7071068 0.7071068 1.4142136 16.2634560 16.9705627 5.6568542 -0.7071068 -3.5355339
[10] -2.8284271 -0.7071068 2.1213203 -0.7071068 -4.9497475 -7.0710678 -9.1923882 -6.3639610 -3.5355339
[19] -2.1213203 0.0000000 -2.1213203 1.4142136 0.7071068 -0.7071068 -0.7071068 0.7071068 0.0000000
[28] 0.7071068 0.0000000 -0.7071068 0.0000000 0.0000000

```

FIGURE 9.4 – Coefficients d'ondelettes correspondant dans l'ordre aux fonctions de base de la figure 9.3 donnée par la fonction `dwt()` du package `waveslim` pour une forme d'onde LiDAR.

1.4 Reconstruction d'un signal de forme d'onde à l'aide des ondelettes de Haar

Nous présentons à présent à titre d'exemple la reconstruction d'une forme d'onde Lidar du Golfe du Morbihan avec le capteur APD en fonction des différents niveaux regroupant des fonctions de base avec des niveaux de résolution de plus en plus fins. Plus on ajoute de niveaux et plus on est précis dans la reconstruction du signal. Notons que le premier niveau permet d'approcher le signal avec **une résolution très grossière**. Les niveaux supérieurs sont ce que l'on appelle "**les détails**" construit avec les ondelettes mères. Plus on garde de niveaux⁵ et plus on sera précis dans la reconstruction du signal.

4. L'algorithme de Mallat décrit une transformation en ondelettes discrètes en utilisant l'analyse multirésolution.

5. Les ondelettes peuvent notamment servir à filtrer un signal bruité en ne prenant pas en compte les fonctions de bases du (ou des) niveau(x) de résolution le(s) plus fin(s)

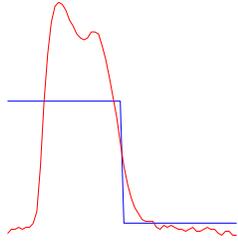


FIGURE 9.5 – Niveau 1 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients)

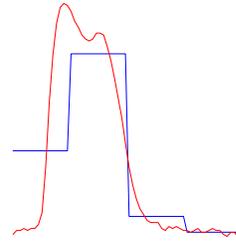


FIGURE 9.6 – Niveau 1+2 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients) et de 2 fonctions d'ondelettes mères (2 coefficients)

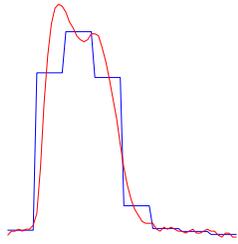


FIGURE 9.7 – Niveau Niveau 1+2+3 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2 et de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients)

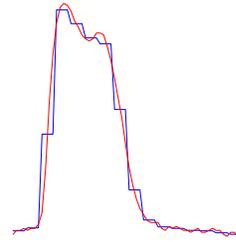


FIGURE 9.8 – Niveau 1+2+3+4 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2 , de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients)

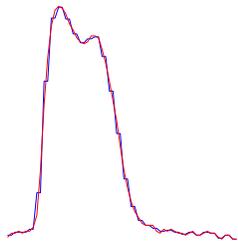


FIGURE 9.9 – Niveau 1+2+3+4+5 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2, de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients) et de 16 fonctions d'ondelettes mère du niveau 5

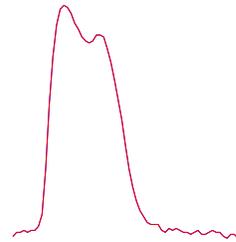


FIGURE 9.10 – Tous les niveaux : 1+2+3+4+5+6 : Combinaison linéaire de 2 fonctions d'échelles pères (2 coefficients), de 2 fonctions d'ondelettes mères (2 coefficients) du niveau 2 , de 4 fonctions d'ondelettes mères du niveau 3 (4 coefficients), de 8 fonctions d'ondelettes mères du niveau 4 (8 coefficients), de 16 fonctions d'ondelettes mère du niveau 5 et de 32 fonctions d'ondelettes mère du niveau 6.

2 Inversion des formes d'onde LiDAR avec l'analyse en ondelettes

Chaque forme d'onde LiDAR peut donc être caractérisée par un nombre de coefficients d'ondelettes. On a vu que pour des formes d'onde LiDAR⁶ constituées de 64 points on avait 64 coefficients en ondelette qui lui étaient associés. Nous allons donc construire un modèle de régression multiple en utilisant **la bathymétrie mesurée sur le terrain comme "variable d'intérêt"** et **les 64 coefficients en ondelettes comme "variables explicatives"**. Pour réduire la dimension du nombre de variables explicatives, nous allons utiliser **"la sélection automatique de variables"⁷** (Stepwise). Dans le cas de la base de données du Golfe du Morbihan⁸, l'algorithme ne retient que 14 variables explicatives sur les 64. On peut donc construire un modèle de régression linéaire multiple avec ces 14 variables explicatives et la bathymétrie mesurée. En utilisant la validation croisée⁹, on peut voir sur la figure 9.11 le résultat du modèle statistique avec la bathymétrie estimée en fonction de la bathymétrie mesurée.

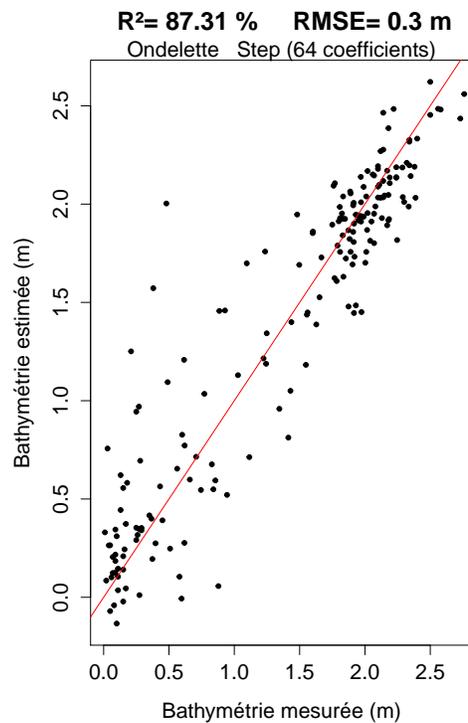


FIGURE 9.11 – Comparaison entre la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction `stepwise()` qui a sélectionnée 14 coefficients d'ondelettes sur les 64 coefficients d'ondelettes (tous les niveaux).

En utilisant le même raisonnement, on peut utiliser la fonction Stepwise sur seulement les coefficients des 5 premiers niveaux (on supprime le niveau 6 les "détails" les plus fins), c'est à dire sur les 32 premiers coefficients¹⁰. on obtient dans ce cas :

6. pour le site du golfe du morbihan (avec le capteur APD)
7. Algorithme permettant à partir d'un nombre de variables explicatives (coefficient en ondelette) et une variable d'intérêt (la bathymétrie) en entrée de nous indiquer seulement les variables qui nous permettent de créer le "meilleur modèle" (cf Annexe I).
8. avec le capteur Avalanche PhotoDiode
9. On crée un modèle avec 80 % de formes d'onde de la base et on test le modèle avec les 20 % de formes d'onde restantes. En répétant le processus 5 fois (5-folds)
10. Niveau 1 : 2 coefficients, niveau 2 : 2 coefficients, niveau 3 : 4 coefficients, niveau 4 : 8 coefficients, niveau 5 : 16 coefficients.

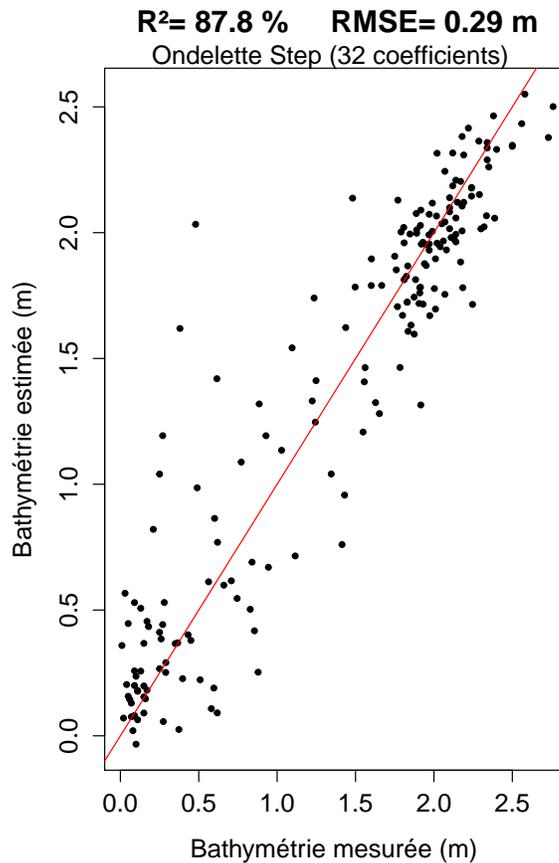


FIGURE 9.12 – Comparaison entre la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction `stepwise()` qui a sélectionnée 11 coefficients d'ondelettes sur les 32 coefficients d'ondelettes des niveaux 1,2,3,4,5 (On ne prend pas en compte le niveau 6)

On applique la fonction `stepwise` :

- d'abord sur toutes les variables explicatives (64 coefficients)
- puis sur les 32 coefficients (on supprime le niveau 6)
- puis sur 16 coefficients (on supprime le niveau 5+6)
- puis sur 8 coefficients (on supprime le niveau 5+6+4)
- et enfin sur 4 coefficients (on supprime le niveau 5+6+4+3)

et on effectue des régressions linéaires multiples, des arbres de décision (CART) et des arbres aléatoires (Random Forest) avec la bathymétrie mesurée et les variables explicatives sélectionnées par la fonction `stepwise()` de **R**. Le tableau 9.1 regroupe les différents modèles statistiques.

Nombre total de coefficients=2+2+4+8+5+16=32

CHAPITRE 9. INVERSION DES FORMES D'ONDE LIDAR AVEC L'ANALYSE EN ONDELETTES

| Méthode utilisée | Modèle statistique | R^2 (%) | RMSE (m) | Nbre de variables descriptifs retenu |
|------------------------------|--------------------|--------------|------------|--------------------------------------|
| Stepwise sur 64 coefficients | RLM | 87.31(88.13) | 0,3(0.29) | 14(15) |
| Stepwise sur 64 coefficients | CART | 79.53(79.72) | 0.38(0.33) | 14(15) |
| Stepwise sur 64 coefficients | RF | 88.2(88.24) | 0.29(0.29) | 14(15) |
| Stepwise sur 32 coefficients | RLM | 87.8(87.39) | 0.29(0.3) | 11(12) |
| Stepwise sur 32 coefficients | CART | 79.16(79.99) | 0.39(0.38) | 11(12) |
| Stepwise sur 32 coefficients | RF | 88.37(88.81) | 0.28(0.28) | 11(12) |
| Stepwise sur 16 coefficients | RLM | 86.74(87.1) | 0.3(0.3) | 6(7) |
| Stepwise sur 16 coefficients | CART | 84.98(81.28) | 0.32(0.37) | 6(7) |
| Stepwise sur 16 coefficients | RF | 89.69(88.82) | 0.27(0.28) | 6(7) |
| Stepwise sur 8 coefficients | RLM | 86.27(86.53) | 0.31(0.31) | 4(5) |
| Stepwise sur 8 coefficients | CART | 81.62(79.49) | 0.36(0.38) | 4(5) |
| Stepwise sur 8 coefficients | RF | 86.24(86.59) | 0.31(0.31) | 4(5) |
| Stepwise sur 4 coefficients | RLM | 85.01(85.98) | 0.32(0.31) | 3(4) |
| Stepwise sur 4 coefficients | CART | 84.9(83.4) | 0.33(0.34) | 3(4) |
| Stepwise sur 4 coefficients | RF | 86.66(85.56) | 0.31(0.32) | 3(4) |

TABLE 9.1 – Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables explicatives des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : RandomForest, stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 64,32,16,8 et 4 variables). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (Partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses.

Chapitre 10

Inversion des formes d'onde LiDAR avec les splines cubiques

Sommaire

| | | |
|----------|---|-----------|
| 1 | Rappel sur les splines cubiques | 82 |
| 1.1 | Principe des splines cubiques | 82 |
| 1.2 | Les coefficients des splines cubiques | 83 |
| 1.3 | Interpolation d'une forme d'onde à l'aide des polynômes cubiques par morceaux | 84 |
| 2 | Inversion des formes d'onde à l'aide des splines cubiques | 85 |

Résumé du chapitre : Dans ce chapitre, je fais un rappel sur les splines cubiques. Cette méthode nous permet d'extraire **les coefficients des polynômes cubiques** représentant les variables explicatives pour la construction des modèles statistiques en utilisant la même approche que j'ai utilisée avec l'Analyse en Composantes Principales et l'analyse en ondelettes. J'explique ensuite toujours sur la base de données du Golfe du Morbihan ¹ comment extraire les coefficients des splines cubiques sur une forme d'onde pour pouvoir dans une seconde étape faire une inversion des formes d'onde en utilisant la fonction `stepwise()` de **R** et construire mes modèles statistiques. Je commente les résultats des modèles d'inversion des formes d'onde LiDAR de toutes les bases de données dans le chapitre 12.

1. avec le capteur Avalanche PhotoDiode

1 Rappel sur les splines cubiques

1.1 Principe des splines cubiques

Une forme d'onde avec un nombre c de points (ou samples) peut-être interpolée par plusieurs polynômes par morceaux appelées **splines**. Les formes d'onde peuvent donc être caractérisées par les coefficients des polynômes des splines (les variables explicatives) qui nous serviront à construire nos modèles statistique (Régression linéaire multiple, arbre de décision et arbres aléatoires).

L'interpolation polynômiale la plus grossière consiste à utiliser une **approximation affine par morceaux** (on relie $(t_v, P(t_v))$ aux points $(t_{v+\Delta t}, P(t_{v+\Delta t}))$ par des segments).

Dans le cadre de ce stage, je me suis intéressé aux **splines cubiques** (approximation cubique par morceaux) représentant **un bon compromis entre l'ordre du degré du polynômes et la régularité de la courbe** obtenue. La figure 10.1 illustre une fonction interpolée par q polynômes cubiques.

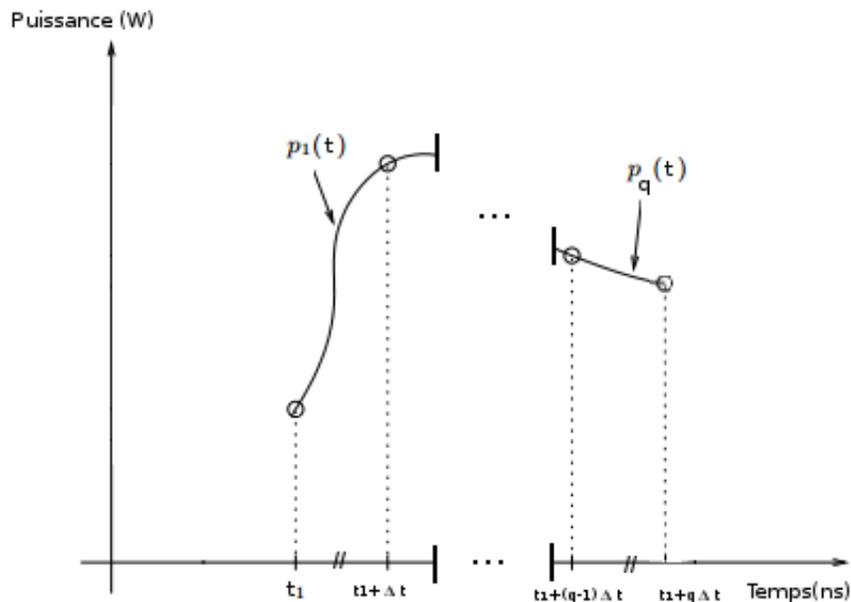


FIGURE 10.1 – Interpolation à l'aide des splines cubiques

Une forme d'onde peut-être interpolée par q polynômes cubiques. On interpole la forme d'onde sur $q + 1$ points (ou samples) avec un pas régulier $\Delta t = \text{Nombre entier inférieur à } \frac{c-1}{q}$. Le polynôme cubique $n^{\circ}v$ de la forme d'onde $n^{\circ}i$ s'écrit :

$$p_v^i(t) = a_v^i(t - (t_1 + (v-1)\Delta t))^3 + b_v^i(t - (t_1 + (v-1)\Delta t))^2 + c_v^i(t - (t_1 + (v-1)\Delta t)) + d_v^i \quad v = 1, \dots, q \quad i = 1, \dots, N \quad (10.1)$$

avec a_v^i, b_v^i, c_v^i et d_v^i les coefficients du polynôme cubique $n^{\circ}v$ de la forme d'onde $n^{\circ}i$. Ces coefficients sont les variables explicatives des formes d'onde.

1.2 Les coefficients des splines cubiques

La fonction `interpSplines()` du package `splines` permet d'obtenir automatiquement les coefficients des polynômes cubiques par morceaux.

Prenons toujours l'exemple d'une forme d'onde de c samples du site du Golfe du Morbihan² et calculons les coefficients des polynômes avec la fonction `interpSpline()`. On choisit au préalable le nombre q de polynômes cubiques par morceaux pour interpoler la forme d'onde LiDAR.

```

1  nombre_polynomes=30 #le nombre q de polynomes cubiques par morceaux pour interpoler une
   forme d'onde LiDAR.
2  nombre_points=nombre_polynomes+1 # q+1 points
3  t=seq(1,64,length=nombre_points) # q+1 points equidistant pour l'interpolation
4  y=signal # Une forme d'onde de 64 points
5  ispl=interpSpline(t, y )

```

```

> ispl
polynomial representation of spline for y ~ t
      constant      linear      quadratic      cubic
1.0  2.3797  6.90565881  0.000000000  0.120766279
3.1  18.0000  8.50339668  0.760827557  0.193026396
5.2  41.0000  14.25261164  1.976893853 -0.609932656
7.3  74.0000  14.48615678 -1.865681881  0.195089932
9.4  98.0000  9.23133268 -0.636615308 -0.062447373
11.5 114.0000  5.73136964 -1.030033757  0.054699559
13.6 122.0000  2.12890303 -0.685426533  0.059608535
15.7 124.0000  0.03873251 -0.309892763  0.030805400
17.8 123.0000 -0.85526165 -0.115818740 -0.074850437
19.9 120.0000 -2.33197164 -0.587376494  0.160616648
22.0 114.0000 -2.67399466  0.424508388 -0.027717656
24.1 110.0000 -1.25776401  0.249887158 -0.049746026
26.2 108.0000 -0.86637787 -0.063512803  0.010742359
28.3 106.0000 -0.99101024  0.004164056  0.114756291
30.4 105.0000  0.54470453  0.727128690 -0.145828424
32.5 108.0000  1.66933498 -0.191590381 -0.071341094
34.6 110.0000 -0.07918730 -0.641039274  0.107253701
36.7 108.0000 -1.35258578  0.034659043 -0.033734611
38.8 105.0000 -1.65332671 -0.177869007 -0.188274657
40.9  99.0000 -4.89125024 -1.363999343  0.354914438
43.0  86.0000 -5.92452947  0.871961616 -0.151586097
45.1  76.0000 -4.26777474 -0.083030795 -0.072509149
47.2  66.0000 -5.57580013 -0.539838435  0.117683594
49.3  53.0000 -6.28616760  0.201568210  0.033693571
51.4  41.0000 -4.99381518  0.413837705 -0.036498477
53.5  32.0000 -3.73857168  0.183897298  0.004320638
55.6  25.0000 -2.90904098  0.211117321  0.019215923
57.7  20.0000 -1.76812156  0.332177637 -0.081184331
59.8  17.0000 -1.44704419 -0.179283651  0.089562003
61.9  14.0000 -1.01513023  0.384956967 -0.061104281

```

FIGURE 10.2 – Coefficients des polynômes cubiques par morceaux pour l'interpolation d'une forme d'onde LiDAR avec la fonction `interpSpline()` du package `splines`

2. avec le capteur APD

1.3 Interpolation d'une forme d'onde à l'aide des polynômes cubiques par morceaux

La figure 10.3 illustre à titre d'exemple l'interpolation d'une forme d'onde réelle de la base de données du Golfe du Morbihan³ en utilisant 5, 10, 20 et 30 polynômes cubiques.

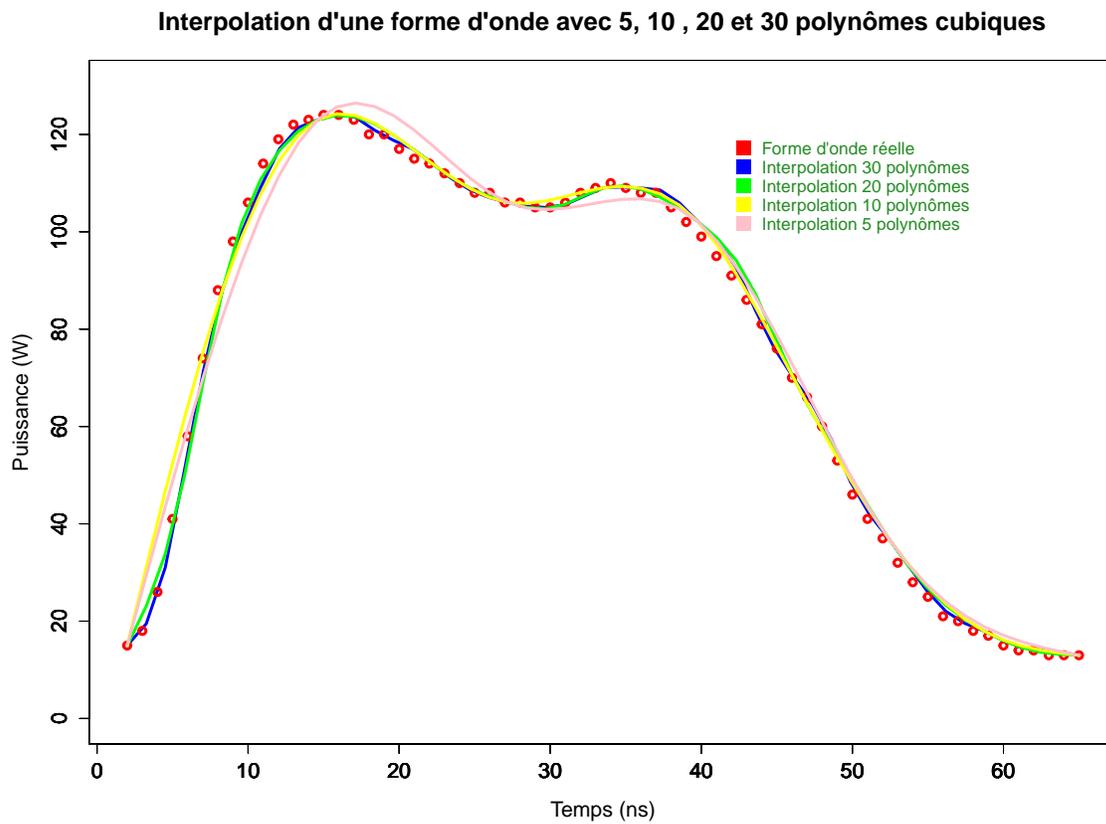


FIGURE 10.3 – Interpolation d'une forme d'onde réelle en utilisant 5, 10, 20 et 30 polynômes cubiques

3. avec le capteur APD

2 Inversion des formes d'onde à l'aide des splines cubiques

Lorsque l'on interpole une forme d'onde Lidar avec **30 polynômes cubiques**, on caractérise la forme d'onde avec 4×30 (120) variables explicatives (soit 4 coefficients a, b, c et d pour chaque polynôme cubique). On utilise la fonction **stepwise** qui permet de sélectionner les variables explicatives (les coefficients des polynômes, soit 4×30 variables explicatives). Dans le cas de la base de données du Golfe du Morbihan⁴, la fonction stepwise sélectionne 17 variables explicatives sur 120 variables explicatives. On peut ainsi construire un modèle de régression linéaire multiple. La figure 10.4 illustre directement la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction stepwise qui a sélectionné 17 variables explicatives.

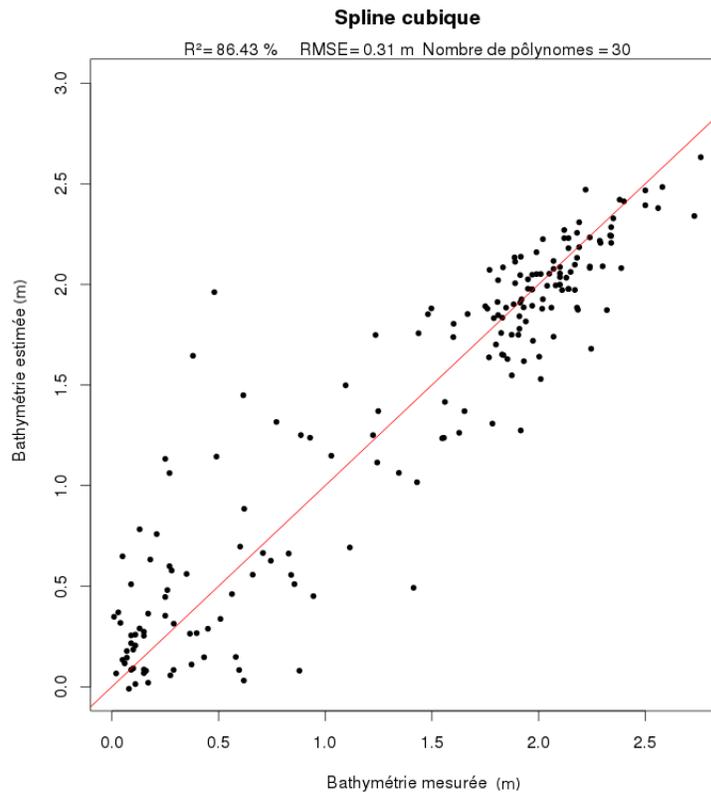


FIGURE 10.4 – Comparaison de la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction stepwise qui a sélectionné 17 variables explicatives sur 120 variables explicatives (les coefficients des splines cubiques).

4. avec le capteur APD

CHAPITRE 10. INVERSION DES FORMES D'ONDE LIDAR AVEC LES SPLINES CUBIQUES

On peut également interpoler les formes d'onde par 20,10 et 5 polynômes cubiques. Les résultats des différents modèles statistiques sont illustrés dans le tableau 11.2.

| Méthode utilisée | Modèle statistique | R^2 (%) | RMSE (m) | Nbre de variables descriptifs retenu |
|--|--------------------|--------------|------------|--------------------------------------|
| Stepwise sur 120 coefficients (30 polynômes) | RLM | 86.43(86.96) | 0.31(0.3) | 17(18) |
| Stepwise sur 120 coefficients (30 polynômes) | CART | 60.01(62.85) | 0.53(0.51) | 17(18) |
| Stepwise sur 120 coefficients (30 polynômes) | RF | 65.17(73.27) | 0.49(0.43) | 17(18) |
| Stepwise sur 80 coefficients (20 polynômes) | RLM | 87.48(85.28) | 0.3(0.32) | 15(16) |
| Stepwise sur 80 coefficients (20 polynômes) | CART | 68.85(73.06) | 0.47(0.43) | 15(16) |
| Stepwise sur 80 coefficients (20 polynômes) | RF | 78.44(78.57) | 0.39(0.39) | 15(16) |
| Stepwise sur 40 coefficients (10 polynômes) | RLM | 85(85.85) | 0.32(0.31) | 7(8) |
| Stepwise sur 40 coefficients (10 polynômes) | CART | 64.3(66.55) | 0.5(0.48) | 7(8) |
| Stepwise sur 40 coefficients (10 polynômes) | RF | 76.76(79.56) | 0.4(0.38) | 7(8) |
| Stepwise sur 20 coefficients (5 polynômes) | RLM | 85.37(85.37) | 0.32(0.32) | 3(4) |
| Stepwise sur 20 coefficients (5 polynômes) | CART | 69.75(65.46) | 0.46(0.49) | 3(4) |
| Stepwise sur 20 coefficients (5 polynômes) | RF | 73.45(79) | 0.43(0.38) | 3(4) |

TABLE 10.1 – Tableau qui regroupe le coefficient de détermination R^2 , le RMSE et le nombre de variables explicatives des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : RandomForest ; Stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 120, 80, 40 et 20 variables explicatives). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (Partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses.

Chapitre 11

Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques

Sommaire

| | | |
|----------|---|-----------|
| 1 | Rappel sur les B-Splines cubiques | 88 |
| 1.1 | Principe des B-Splines cubiques | 88 |
| 1.2 | Les fonctions de base | 89 |
| 1.3 | Les coefficients des B-Splines cubiques | 90 |
| 2 | Reconstruction d'une forme d'onde réelle avec les B-Splines cubiques | 91 |
| 3 | Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques | 92 |

Résumé du chapitre : Dans ce chapitre, je fais un rappel sur les B-Splines cubiques. Cette méthode nous permet d'extraire **les coefficients des B-Splines cubiques** représentant les variables explicatives pour la construction des modèles statistiques en utilisant la même approche qui a été utilisée avec l'Analyse en Composantes Principales, l'analyse en ondelettes et les splines cubiques. J'explique ensuite toujours sur la base de données du Golfe du Morbihan ¹ comment calculer les coefficients des B-Splines cubiques avec le package **fda** sur une forme d'onde pour pouvoir dans une seconde étape faire une inversion des formes d'onde en utilisant la fonction `stepwise()` de **R** et construire mes modèles statistiques. Je commente les résultats des modèles d'inversion des formes d'onde LiDAR de toutes les bases de données dans le chapitre 12.

1. avec le capteur Avalanche PhotoDiode

1 Rappel sur les B-Splines cubiques

1.1 Principe des B-Splines cubiques

A la différence des splines cubiques (qui interpolent une forme d'onde avec une somme de polynômes cubiques par morceaux), les B-Splines (Basis-Splines) permettent d'approcher les formes d'onde LiDAR avec une combinaison linéaire de fonctions de base qui sont des polynômes cubiques. On choisit au préalable le nombre p de fonctions de base que l'on va utiliser pour approcher la forme d'onde. Le but est d'extraire **les coefficients des fonctions de base** représentant les variables explicatives qui serviront à construire nos modèles statistiques. Notons que nous avons autant de fonctions de base que de coefficients. La forme d'onde $n^{\circ}i$ peut-être approximée par la formule suivante :

$$P^i(t) = \sum_{j=1}^p C_j^i B_j^i(t) \quad (11.1)$$

avec C_j^i le coefficient de la fonction de base $n^{\circ}j$ de la forme d'onde $n^{\circ}i$, $B_j^i(t)$ la fonction de base $n^{\circ}j$ de la forme d'onde $n^{\circ}i$, $P^i(t)$ est la fonction qui approche la forme d'onde $n^{\circ}i$. Pour plus d'informations sur les B-Splines cubiques cf. [48] et [2].

1.2 Les fonctions de base

Les fonctions de base B-Splines (Basis-splines) se calculent à l'aide d'un l'algorithme récursif : **l'algorithme de Cox-Boor**. La figure 1.1 illustre $p=5, 10, 20$ et 30 fonctions de base qui permettent d'approcher une forme d'onde LiDAR.

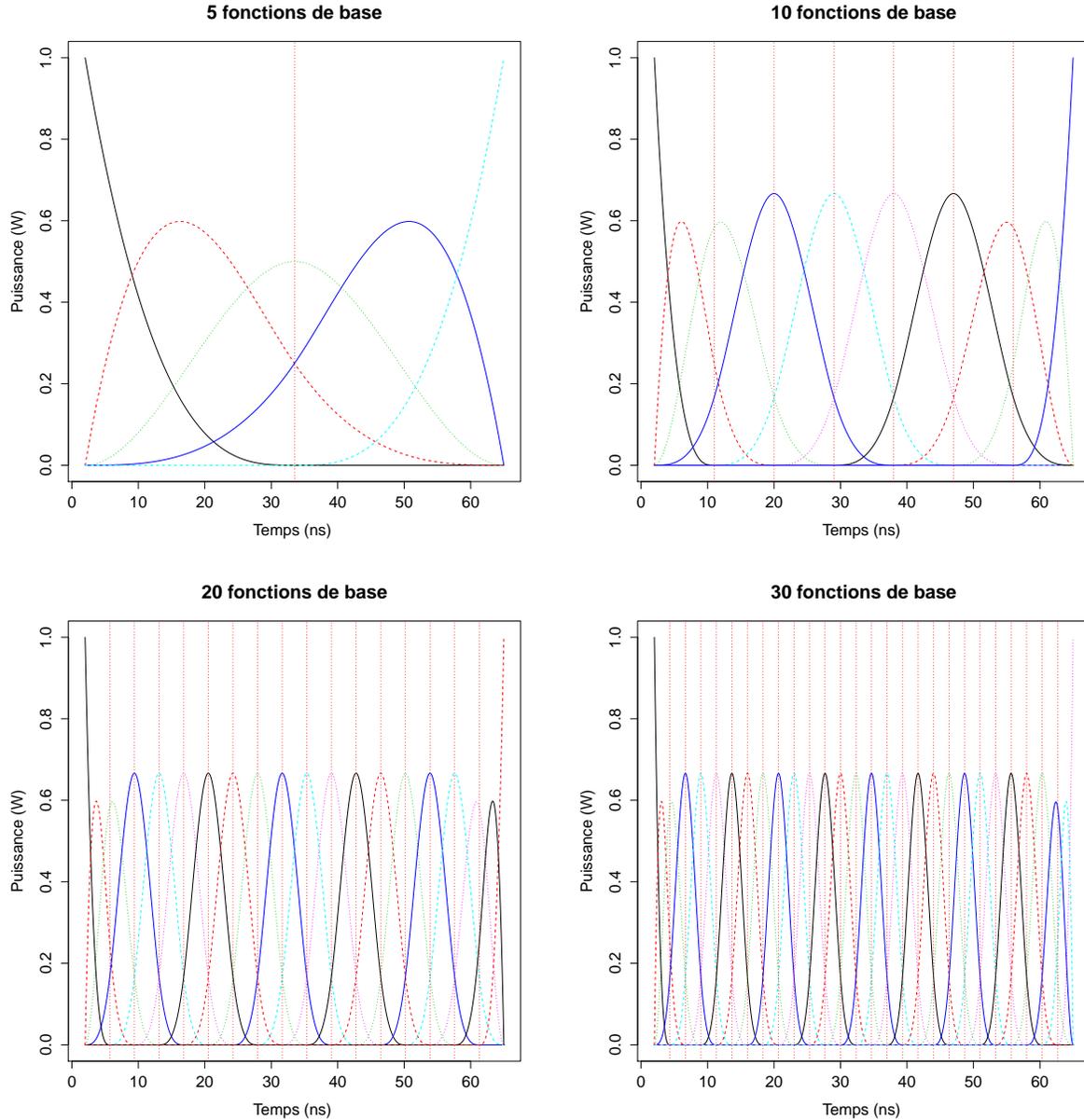


TABLE 11.1 – Base de 5,10, 20 et 30 B-Splines cubiques uniformes. Une fonction de base est un polynôme de degré 3.

1.3 Les coefficients des B-Splines cubiques

Une forme d'onde peut-être approchée par la combinaison linéaire de p fonctions de base. Je présente ici comment calculer les coefficients² des B-Splines pour $p=5$ en utilisant le package `fda` sur une forme d'onde du Golfe du Morbihan³ pour pouvoir ensuite calculer dans un second temps les coefficients des B-Splines cubiques sur toutes les formes d'onde de la base du Golfe du Morbihan et construire mes modèles statistiques.

```

1
2 i=10                # Numero de la forme d'onde
3 n=5                 # Choix du nombre de fonctions de base
4 Signal=tableau[i,1:64] # Forme d'onde
5 y=t(Signal)        # Transposee
6
7 # c(1,64) : Debut et fin Signal
8 # norder=4 : ordre du polynome (degre polynome+1=ordre polynome)
9
10 basis = create.bspline.basis(c(1,64), nbasis=n, norder=4)
11
12 # lambda est un parametre de lissage
13 fdParobj = fdPar(fdobj=basis, Lfdobj=2, lambda=0)
14
15 plfd = smooth.basis(1:64,y,fdParobj=fdParobj)$fd # Approximation de la forme d'onde
16
17 print(plfd$coefs) # Affiche les variables explicatives (coefficients des B-Splines)

```

```

> print(plfd$coefs)
      [,1]
bspl4.1 -14.288800
bspl4.2  313.683014
bspl4.3 -109.463954
bspl4.4   43.567383
bspl4.5   1.309859

```

FIGURE 11.1 – Coefficients des B-Splines associés à la figure 1.1 pour le cas des 5 fonctions de base

2. Pour plus d'informations, cf. [2]

3. avec le capteur APD

2 Reconstruction d'une forme d'onde réelle avec les B-Splines cubiques

La figure 11.2 illustre à titre d'exemple l'approximation d'une forme d'onde réelle du Golfe du Morbihan⁴ avec 5, 10, 20 et 30 fonctions de base.

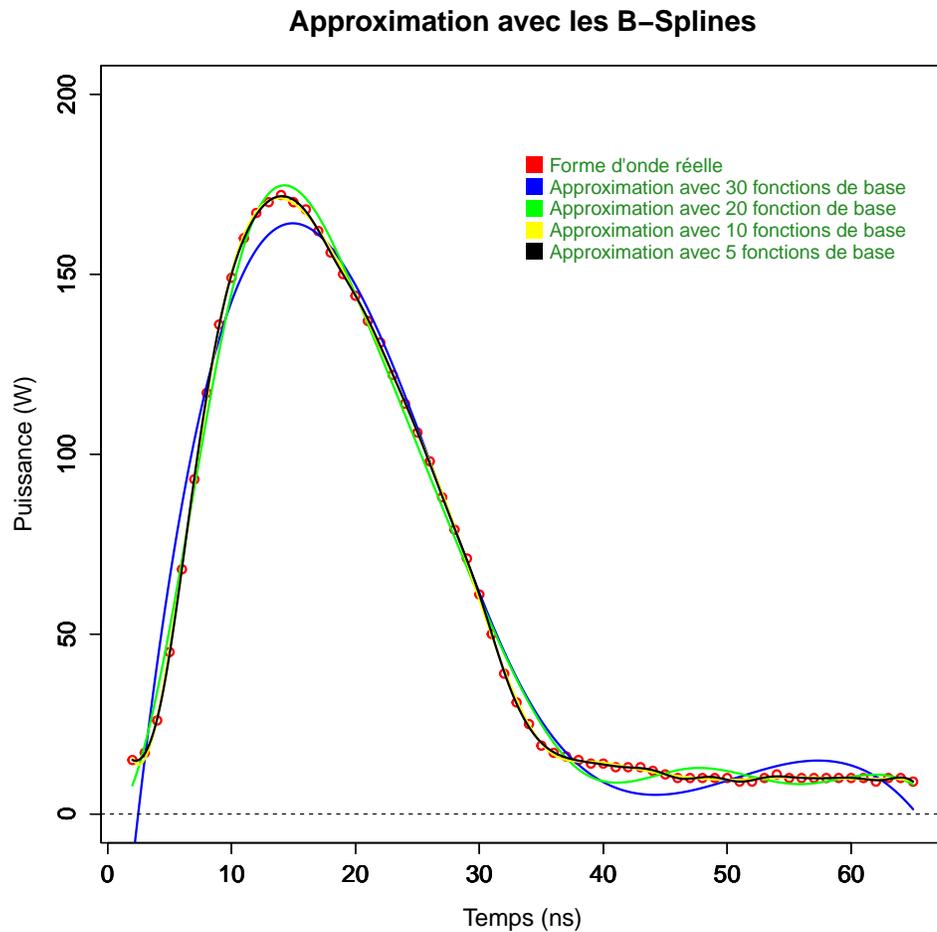


FIGURE 11.2 – Approximation d'une forme d'onde réelle du Golfe du Morbihan avec les B-Splines cubiques.

4. avec le capteur Avalanche PhotoDiode

3 Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques

On peut donc approximer chaque forme d'onde LiDAR avec p fonctions de base. On caractérise une forme d'onde avec p variables explicatives (les p coefficients des fonctions de base). Dans le cas de la base de données du Golfe du Morbihan, on approche chaque forme d'onde avec $p=5$ fonctions de base, ce qui nous permet d'avoir $p=5$ variables explicatives. J'utilise ensuite la fonction `stepwise()` qui sélectionne 3 variables explicatives sur 5. On peut ainsi construire un modèle de régression linéaire multiple en utilisant toujours notre approche statistique. La figure 11.3 illustre directement la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction `stepwise` qui a sélectionné 3 variables explicatives.

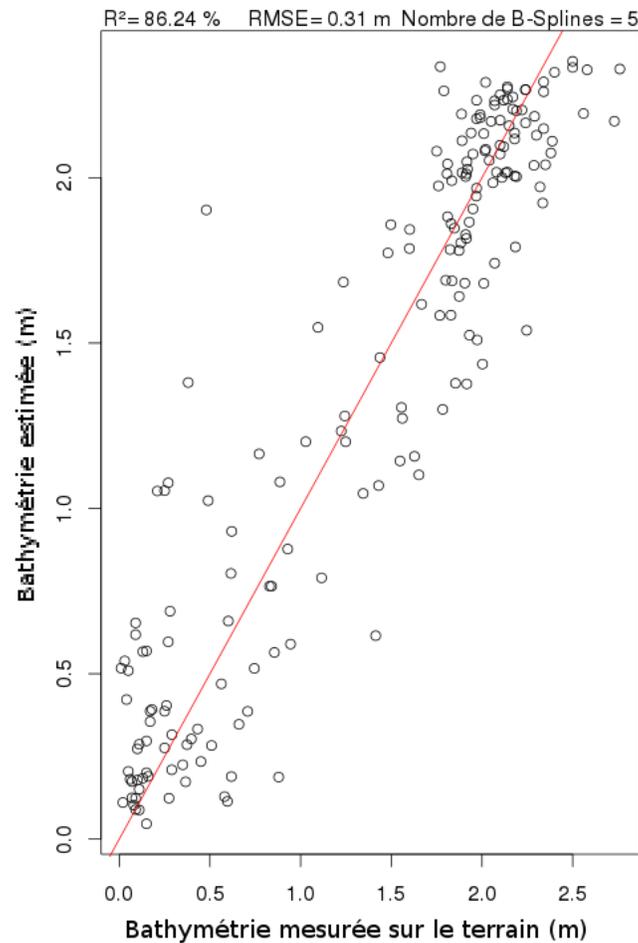


FIGURE 11.3 – Comparaison de la bathymétrie estimée en fonction de la bathymétrie mesurée en utilisant la fonction `stepwise` qui a sélectionné 3 variables explicatives sur 5 variables explicatives (les coefficients des B-Splines cubiques).

CHAPITRE 11. INVERSION DES FORMES D'ONDE LIDAR À L'AIDE DES B-SPLINES CUBIQUES

Les résultats des différents modèles statistiques sont illustrés dans le tableau 11.2 pour $p=5, 10, 20$ et 30 .

| Méthode utilisée | Modèle statistique | R^2 (%) | RMSE (m) | Nbre de variables descriptifs |
|--|--------------------|--------------|------------|-------------------------------|
| Stepwise sur 5 variables explicatives | RLM | 86.24(84.83) | 0.31(0.33) | 3(4) |
| Stepwise sur 5 variables explicatives | CART | 83.9(81.13) | 0.34(0.36) | 3(4) |
| Stepwise sur 5 variables explicatives | RF | 86.3(87.63) | 0.31(0.29) | 3(4) |
| Stepwise sur 10 variables explicatives | RLM | 86.59(86.78) | 0.31(0.3) | 4(5) |
| Stepwise sur 10 variables explicatives | CART | 78.82(82.93) | 0.39(0.35) | 4(5) |
| Stepwise sur 10 variables explicatives | RF | 84.43(83.75) | 0.33(0.34) | 4(5) |
| Stepwise sur 20 variables explicatives | RLM | 87.71(87.95) | 0.29(0.29) | 4(5) |
| Stepwise sur 20 variables explicatives | CART | 81.28(78.83) | 0.36(0.39) | 4(5) |
| Stepwise sur 20 variables explicatives | RF | 85.61(86.52) | 0.32(0.31) | 4(5) |
| Stepwise sur 30 variables explicatives | RLM | 86.4(87.28) | 0.31(0.3) | 11(12) |
| Stepwise sur 30 variables explicatives | CART | 81.52(81.09) | 0.36(0.37) | 11(12) |
| Stepwise sur 30 variables explicatives | RF | 88.46(87.26) | 0.29(0.3) | 11(12) |

TABLE 11.2 – Coefficient de détermination R^2 et le RMSE des différents modèles statistiques (RLM : Régression linéaire multiple, CART : Arbres décisionnels, RF : Random Forest ; Stepwise : algorithme qui sélectionne les variables descriptives qui contribuent le plus à la régression parmi les 5,10,20 et 30 variables explicatives (5, 10, 20 et 30 fonctions de base). On ajoute à tous les modèles statistiques la variable explicative W_{ext} (partie utile des formes d'onde). Les résultats des modèles statistiques avec W_{ext} sont entre parenthèses.

Quatrième partie

Résultats obtenus et Conclusion

Chapitre 12

Résultats et Analyses

Sommaire

| | | |
|----------|---|------------|
| 1 | Résultats pour les données LiDAR bathymétriques du Golfe du Morbihan . . | 95 |
| 1.1 | Résultats pour la base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode (APD) | 95 |
| 1.2 | Résultats pour la base de données du Golfe du Morbihan avec le capteur Photo-Multiplicateur (PMT) | 97 |
| 2 | Résultats pour les données LiDAR acquises avec le simulateur Wa-LID . . . | 100 |
| 2.1 | Résultats pour les données LiDAR avec une eau claire | 100 |
| 2.2 | Résultats pour les données LiDAR avec une eau turbide | 103 |
| 3 | Résultats pour les données LiDAR altimétriques du Brésil | 106 |

Résumé du chapitre : Dans ce chapitre, je présente les résultats de chaque base de données sous forme de tableau en indiquant le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives pour les différents modèles statistiques (Régression linéaire multiple, CART¹ et Random Forest²) avec les méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) qui m'ont permis de calculer les variables explicatives des formes d'onde LiDAR. J'expose ensuite sous forme de diagrammes à barres pour chaque base de données les résultats des meilleurs modèles statistiques³ (Régression linéaire multiple, CART et Random Forest) qui contiennent moins de 5 variables explicatives.

1 Résultats pour les données LiDAR bathymétriques du Golfe du Morbihan

1.1 Résultats pour la base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode (APD)

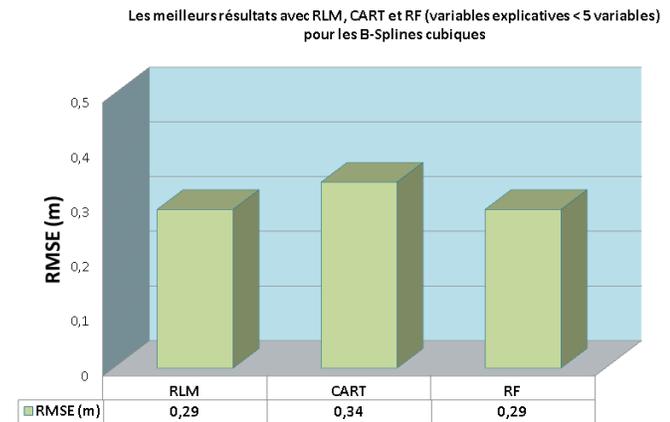
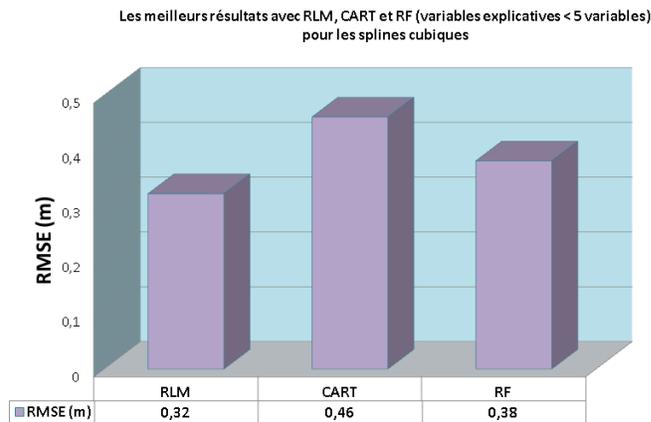
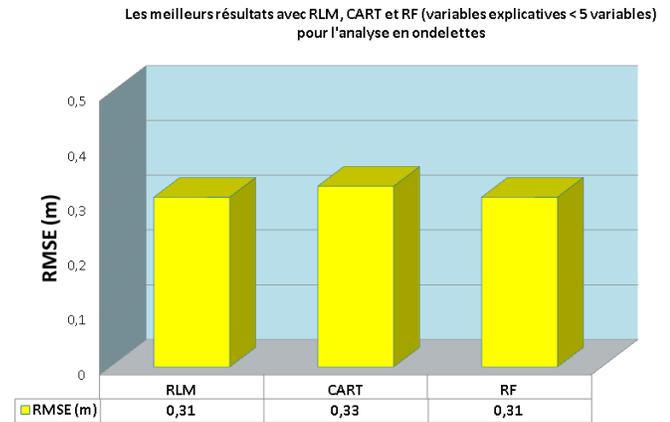
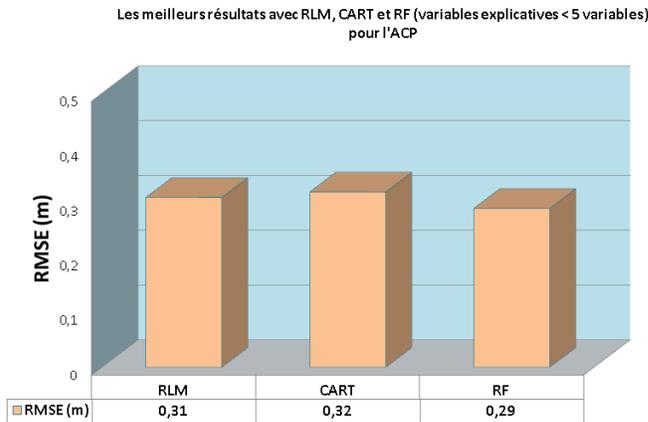
La figure 12.1 fournit le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) pour les données LiDAR du Golfe du Morbihan avec le capteur APD.

1. arbre de décision
2. Les arbres aléatoires
3. RMSE faible

| | R^2 (en %) | RMSE (en m) | Nombre de variables |
|--------------------------------|---|--------------|---------------------|
| ACP | Régression linéaire PCI | 84.7 | 0.33 |
| | Régression linéaire PCI+PC2 | 85.65 | 0.32 |
| | Régression linéaire PCI+PC2+PC3 | 85.83 | 0.31 |
| | Régression linéaire PCI+weat | 84.97 | 0.32 |
| | Régression linéaire PCI+PC2+weat | 85.93 | 0.31 |
| | Régression linéaire PCI+PC2+PC3+weat | 85.44 | 0.32 |
| | CART PCI | 80.94 | 0.37 |
| | CART PCI+PC2 | 83.88 | 0.34 |
| | CART PCI+PC2+PC3 | 84.58 | 0.33 |
| | CART PCI+weat | 81.81 | 0.36 |
| | CART PCI+PC2+weat | 85.59 | 0.32 |
| | CART PCI+PC2+PC3+weat | 83.37 | 0.34 |
| | RF PCI | 79.83 | 0.38 |
| | RF PCI+PC2 | 88.15 | 0.29 |
| | RF PCI+PC2+PC3 | 85.54 | 0.32 |
| | RF PCI+weat | 80.43 | 0.37 |
| | RF PCI+PC2+weat | 84.44 | 0.33 |
| | RF PCI+PC2+PC3+weat | 87.38(86.81) | 0.30(3) |
| | Régression linéaire + Stepwise | 83.16(84.07) | 0.34(0.33) |
| | CART + stepwise | 87.49(86.19) | 0.32(0.32) |
| Ondelettes | Régression linéaire + Stepwise (256 coefficients) | | |
| | CART + Stepwise (256 coefficients) | | |
| | RF + Stepwise (256 coefficients) | | |
| | Régression linéaire + Stepwise (128 coefficients) | | |
| | CART + Stepwise (128 coefficients) | | |
| | RF + Stepwise (128 coefficients) | | |
| | RF (128 coefficients) | | |
| | Régression linéaire + Stepwise (64 coefficients) | 87.31(85.15) | 0.30(29) |
| | CART + Stepwise (64 coefficients) | 79.53(79.72) | 0.38(0.38) |
| | RF + Stepwise (64 coefficients) | 88.2(88.24) | 0.29(0.29) |
| | Régression linéaire + Stepwise (32 coefficients) | 87.8(87.39) | 0.29(0.3) |
| | CART + Stepwise (32 coefficients) | 79.16(79.99) | 0.39(0.38) |
| | RF + Stepwise (32 coefficients) | 88.37(88.81) | 0.28(0.28) |
| | Régression linéaire + Stepwise (16 coefficients) | 86.74(87.1) | 0.30(3) |
| | CART (16 coefficients) | 84.98(81.28) | 0.32(0.37) |
| | RF + Stepwise (16 coefficients) | 89.69(88.82) | 0.27(0.28) |
| | Régression linéaire + Stepwise (8 coefficients) | 86.27(86.53) | 0.31(0.31) |
| | CART + Stepwise (8 coefficients) | 81.62(79.49) | 0.36(0.38) |
| | RF + Stepwise (8 coefficients) | 86.24(86.59) | 0.31(0.31) |
| | Régression linéaire + Stepwise (4 coefficients) | 85.01(85.98) | 0.32(0.31) |
| CART (4 coefficients) | 84.9(83.41) | 0.33(0.34) | |
| RF + Stepwise (4 coefficients) | 86.66(85.56) | 0.31(0.32) | |
| Splines cubiques | Régression + Stepwise 30 polynômes | 86.43(86.96) | 0.31(0.3) |
| | Régression + Stepwise 20 polynômes | 87.48(85.28) | 0.30(32) |
| | Régression + Stepwise 10 polynômes | 85(85.85) | 0.32(0.31) |
| | Régression + Stepwise 5 polynômes | 85.37(85.7) | 0.32(0.32) |
| | CART + Stepwise 30 polynômes | 60.01(62.85) | 0.53(0.51) |
| | CART + Stepwise 20 polynômes | 68.85(73.06) | 0.47(0.43) |
| | CART + Stepwise 10 polynômes | 64.3(66.55) | 0.50(48) |
| | CART + Stepwise 5 polynômes | 69.75(65.46) | 0.46(0.49) |
| | RF + Stepwise 30 polynômes | 65.17(73.27) | 0.49(0.43) |
| | RF + Stepwise 20 polynômes | 78.44(78.57) | 0.39(0.39) |
| | RF + Stepwise 10 polynômes | 76.76(75.56) | 0.40(38) |
| | RF + Stepwise 5 polynômes | 73.45(79) | 0.43(0.38) |
| | Régression linéaire + Stepwise 5 B-spline | 86.24(84.83) | 0.31(0.33) |
| | CART + Stepwise 5 B-spline | 83.9(81.13) | 0.34(0.36) |
| | RF + Stepwise 5 B-spline | 86.3(87.63) | 0.31(0.29) |
| B-splines | Régression linéaire + Stepwise 10 B-spline | 86.59(86.78) | 0.31(0.3) |
| | CART + Stepwise 10 B-spline | 78.82(82.93) | 0.39(0.35) |
| | RF + Stepwise 10 B-spline | 84.43(83.75) | 0.33(0.34) |
| | Régression linéaire + Stepwise 20 B-spline | 87.71(87.95) | 0.29(0.29) |
| | CART + Stepwise 20 B-spline | 81.28(78.33) | 0.36(0.39) |
| | RF + Stepwise 20 B-spline | 83.61(86.52) | 0.32(0.31) |
| | Régression linéaire + Stepwise 30 B-spline | 86.43(87.28) | 0.31(0.3) |
| | CART + Stepwise 30 B-spline | 81.52(81.09) | 0.36(0.37) |
| | RF + Stepwise 30 B-spline | 88.46(87.26) | 0.29(0.3) |

TABLE 12.1 – Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques). Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Base de données du Golfe du Morbihan avec le capteur Avalanche PhotoDiode.

A partir du tableau 12.1, on choisit les meilleurs modèles statistiques (Régression linéaire multiple, CART et Random Forest) parmi les quatre méthodes (l'Analyse en Composantes Principales, l'analyse en ondelettes, les splines cubiques et les B-Splines) qui contiennent moins de 5 variables explicatives (il est plus pertinent de choisir des modèles statistiques avec peu de variables explicatives). J'expose sous forme de diagrammes à barres les meilleurs modèles statistiques⁴ (construits avec moins de 5 variables explicatives) pour chacune des méthodes utilisées.



Si on compare le RMSE des meilleurs modèles statistiques pour chacune des méthodes, on constate :

- qu'on obtient des résultats proches pour la régression linéaire multiple et Random Forest.
- et que CART est moins précis (plus particulièrement pour les splines cubiques).

1.2 Résultats pour la base de données du Golfe du Morbihan avec le capteur PhotoMultiplicateur (PMT)

La figure 12.2 fournit le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) pour les données LiDAR du Golfe du Morbihan avec le capteur PMT.

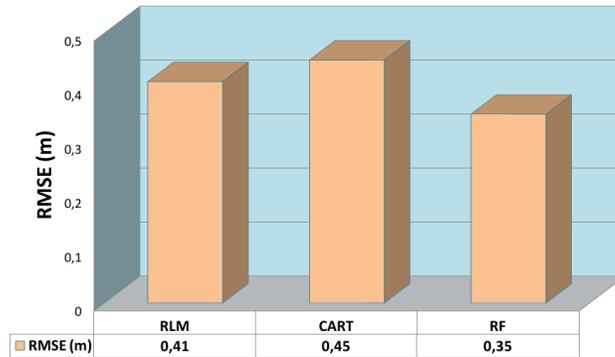
4. Régression linéaire multiple, CART et Random Forest

| | | Morbihan (capteur 2) PMT (0.3m) [189 wf -128] | | Nombre de variables | |
|---|--|---|-------------|---------------------|--|
| | | R ² (en %) | RMSE (en m) | | |
| ACP | Régression linéaire PCI | 65.36 | 0.49 | 1 | |
| | Régression linéaire PCI-PC2 | 67.15 | 0.48 | 2 | |
| | Régression linéaire PCI-PC2-PC3 | 75.71 | 0.41 | 3 | |
| | Régression linéaire PCI-weat | 66.51 | 0.48 | 2 | |
| | Régression linéaire PCI-PC2-weat | 67.49 | 0.48 | 3 | |
| | Régression linéaire PCI-PC2-PC3-weat | 74.67 | 0.42 | 4 | |
| | CART PCI | 62.66 | 0.51 | 1 | |
| | CART PCI-PC2 | 63.87 | 0.51 | 2 | |
| | CART PCI-PC2-PC3 | 70.51 | 0.46 | 3 | |
| | CART PCI-weat | 65.03 | 0.5 | 2 | |
| | CART PCI-PC2-weat | 63.85 | 0.51 | 3 | |
| | CART PCI-PC2-PC3-weat | 71.01 | 0.45 | 4 | |
| Ondelettes | RF PCI | 65.1 | 0.5 | 1 | |
| | RF PCI-PC2 | 69.85 | 0.46 | 2 | |
| | RF PCI-PC2-PC3 | 83.66 | 0.35 | 3 | |
| | RF PCI-weat | 66.24 | 0.49 | 2 | |
| | RF PCI-PC2-weat | 76.02 | 0.41 | 3 | |
| | RF PCI-PC2-PC3-weat | 81.4 | 0.36 | 4 | |
| | Régression linéaire + Stepwise | 85.17(82.94) | 0.32(0.35) | 11(12) | |
| | CART + stepwise | 71.95(65.76) | 0.45(0.5) | 11(12) | |
| | RF + stepwise | 83.56(78.44) | 0.38(0.33) | 11(12) | |
| | Régression linéaire + Stepwise (256 coefficients) | | | | |
| | CART + Stepwise (256 coefficients) | | | | |
| | RF + Stepwise (256 coefficients) | | | | |
| Régression linéaire + Stepwise (128 coefficients) | 87.53(84.83) | 0.30(33) | 76(77) | | |
| CART + Stepwise (128 coefficients) | 76.84(66.81) | 0.40(49) | 76(77) | | |
| RF (128 coefficients) | 86.04(83.99) | 0.31(0.34) | 76(77) | | |
| Régression linéaire + Stepwise (64 coefficients) | 86.61(86.23) | 0.31(0.31) | 23(24) | | |
| CART + Stepwise (64 coefficients) | 75.26(76.4) | 0.42(0.41) | 23(24) | | |
| RF + Stepwise (64 coefficients) | 86.43(83.11) | 0.31(0.34) | 23(24) | | |
| Régression linéaire + Stepwise (32 coefficients) | 85.56(84.47) | 0.32(0.33) | 17(18) | | |
| CART + Stepwise (32 coefficients) | 76.61(74.7) | 0.41(0.43) | 17(18) | | |
| RF + Stepwise (32 coefficients) | 86.06(84.28) | 0.31(0.33) | 17(18) | | |
| Régression linéaire + Stepwise (16 coefficients) | 84.96(84.91) | 0.32(0.32) | 9(10) | | |
| CART (16 coefficients) | 80.11(77.22) | 0.37(0.4) | 9(10) | | |
| RF + Stepwise (16 coefficients) | 85.94(83.73) | 0.32(0.34) | 9(10) | | |
| Régression linéaire + Stepwise (8 coefficients) | 84.14(83.58) | 0.33(0.34) | 5(6) | | |
| CART + Stepwise (8 coefficients) | 70.11(64.48) | 0.46(0.48) | 5(6) | | |
| RF + Stepwise (8 coefficients) | 80.78(79.52) | 0.37(0.38) | 5(6) | | |
| Régression linéaire + Stepwise (4 coefficients) | 70.95(69.28) | 0.49(0.46) | 4(5) | | |
| CART (4 coefficients) | 63.63(63.1) | 0.51(0.47) | 4(5) | | |
| RF + Stepwise (4 coefficients) | 76.67(76.38) | 0.4(0.41) | 4(5) | | |
| Régression + Stepwise 30 polynomes | 83.95(83.63) | 0.34(0.34) | 26(27) | | |
| Régression + Stepwise 20 polynomes | 86.36(86.94) | 0.31(0.31) | 10(11) | | |
| Régression + Stepwise 10 polynomes | 80.65(81.99) | 0.37(0.35) | 11(12) | | |
| Régression + Stepwise 5 polynomes | 75.32(76.35) | 0.41(0.41) | 5(6) | | |
| CART + Stepwise 30 polynomes | 10.78(86.07) | 0.79(0.49) | 26(27) | | |
| CART + Stepwise 20 polynomes | 36.99(66.66) | 0.66(0.48) | 10(11) | | |
| CART + Stepwise 10 polynomes | 66.25(63.08) | 0.49(0.51) | 11(12) | | |
| CART + Stepwise 5 polynomes | 74.51(68.98) | 0.42(0.47) | 5(6) | | |
| RF + Stepwise 30 polynomes | 59.22(73.63) | 0.42(0.43) | 26(27) | | |
| RF + Stepwise 20 polynomes | 63.43(74.32) | 0.5(0.42) | 10(11) | | |
| RF + Stepwise 10 polynomes | 76.80(78.73) | 0.4(0.39) | 11(12) | | |
| RF + Stepwise 5 polynomes | 78.84(81.2) | 0.38(0.36) | 5(6) | | |
| Régression linéaire+Stepwise 5B-spline | 76.11(76.44) | 0.41(0.41) | 4(5) | | |
| CART + Stepwise 5 B-spline | 71.56(71.82) | 0.45(0.45) | 4(5) | | |
| RF + Stepwise 5 B-spline | 80.93(81.22) | 0.37(0.36) | 4(5) | | |
| Régression linéaire+Stepwise 10 B-spline | 81.88(76.55) | 0.36(0.4) | 6(7) | | |
| CART + Stepwise 10 B-spline | 77.32(78.58) | 0.41(0.39) | 6(7) | | |
| RF + Stepwise 10 B-spline | 85.25(80.61) | 0.32(0.37) | 6(7) | | |
| Régression linéaire+Stepwise 20 B-spline | 85.91(83.78) | 0.32(0.34) | 10(11) | | |
| CART + Stepwise 20 B-spline | 73.1(73.85) | 0.44(0.41) | 10(11) | | |
| RF + Stepwise 20 B-spline | 85.4(84.71) | 0.32(0.33) | 10(11) | | |
| Régression linéaire+Stepwise 30 B-spline | 85.32(80.47) | 0.31(0.37) | 9(10) | | |
| CART + Stepwise 30 B-spline | 78.33(74.18) | 0.39(0.43) | 9(10) | | |
| RF + Stepwise 30 B-spline | 86.6(83.84) | 0.31(0.34) | 9(10) | | |

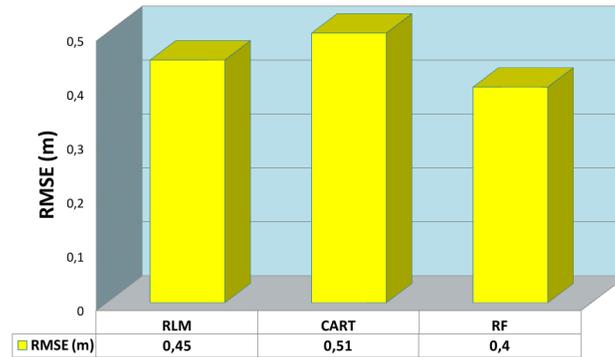
TABLE 12.2 – Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d’onde) - Base de données du Golfe du Morbihan avec le capteur PhotoMultiplicateur Tube.

A partir du tableau 12.2, on choisit les meilleurs modèles statistiques (Régression linéaire multiple, CART et Random Forest) parmi trois méthodes (l'Analyse en Composantes Principales, l'analyse en ondelettes et les B-Splines cubiques) qui contiennent moins de 5 variables explicatives (Les splines cubiques contiennent plus de 4 variables explicatives). J'expose sous forme de diagrammes à barres les meilleurs modèles statistiques⁵ pour chacune des méthodes utilisées.

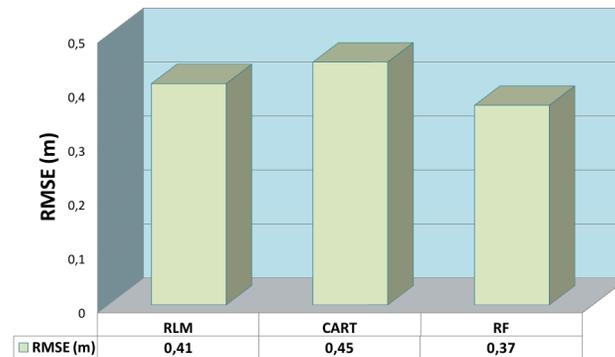
Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour l'ACP



Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour l'analyse en ondelettes



Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour les B-Splines



5. Régression linéaire multiple, CART et Random Forest

Si on compare le RMSE des meilleurs modèles statistiques pour chacune des méthodes, on constate :

- que Random Forest donne de meilleurs résultats suivi de la régression linéaire multiple
- que CART est encore une fois le modèle statistique le moins précis.

2 Résultats pour les données LiDAR acquises avec le simulateur Wa-LID

2.1 Résultats pour les données LiDAR avec une eau claire

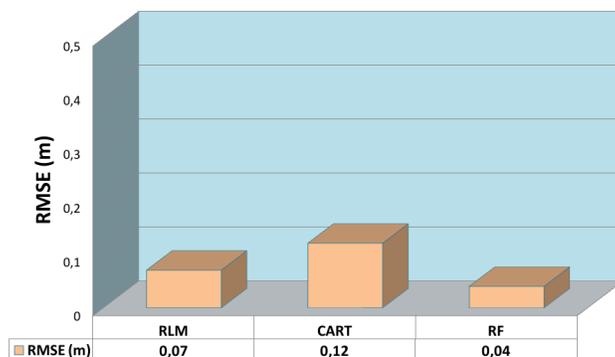
La figure 12.3 fournit le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) pour les données LiDAR simulées avec une eau claire.

| | | Formes ondes simulées $r=0.45$ $\lambda=0.2$ $\theta_{max}=0.00019$ $\sigma_{\theta}=0.01$ $S=2.5$ $C_{th}=0.4$ $\theta_{th}=0.1$ | | |
|--|---|---|-------------|---------------------|
| | | R^2 (en %) | RMSE (en m) | Nombre de variables |
| ACP | Régression linéaire PCI | 89.16 | 0.25 | 1 |
| | Régression linéaire PCHPC2 | 89.66 | 0.24 | 2 |
| | Régression linéaire PCHPC2-PC3 | 97.36 | 0.12 | 3 |
| | Régression linéaire PCHPC2-weit | 98.38 | 0.08 | 2 |
| | Régression linéaire PCHPC2-PC3-weit | 99.21 | 0.07 | 3 |
| | Régression linéaire PCHPC2-PC3-weit | 99.25 | 0.07 | 4 |
| | CART PCI | 95.05 | 0.17 | 1 |
| | CART PCHPC2 | 97.48 | 0.12 | 2 |
| | CART PCHPC2-PC3 | 97.19 | 0.13 | 3 |
| | CART PCHPC2-weit | 97.45 | 0.12 | 2 |
| | CART PCHPC2-PC3 | 97.53 | 0.12 | 3 |
| | CART PCHPC2-PC3-weit | 97.35 | 0.12 | 4 |
| | RF PCI | 98.23 | 0.1 | 1 |
| | RF PCHPC2 | 99.4 | 0.06 | 2 |
| | RF PCHPC2-PC3 | 99.65 | 0.04 | 3 |
| | RF PCHPC2-weit | 99.59 | 0.05 | 2 |
| | RF PCHPC2-weit | 99.71 | 0.04 | 3 |
| | RF PCHPC2-PC3-weit | 99.72 | 0.04 | 4 |
| | Régression linéaire + Stepwise | 99.93(99.94) | 0.02(0.02) | 25(26) |
| | CART + stepwise | 97.24(97.45) | 0.12(0.12) | 25(26) |
| RF + stepwise | 99.87(99.88) | 0.03(0.03) | 25(26) | |
| Ondelettes | Régression linéaire + Stepwise (256 coefficients) | | | |
| | CART + Stepwise (256 coefficients) | | | |
| | RF + Stepwise (256 coefficients) | | | |
| | Régression linéaire + Stepwise (128 coefficients) | | | |
| | CART + Stepwise (128 coefficients) | | | |
| | RF (128 coefficients) | | | |
| | Régression linéaire + Stepwise (64 coefficients) | 99.94(99.94) | 0.02(0.02) | 37(38) |
| | CART + Stepwise (64 coefficients) | 97.08(96.75) | 0.13(0.14) | 37(38) |
| | RF + Stepwise (64 coefficients) | 99.89(99.92) | 0.03(0.02) | 37(38) |
| | Régression linéaire + Stepwise (32 coefficients) | 99.86(99.88) | 0.03(0.03) | 21(22) |
| | CART + Stepwise (32 coefficients) | 96.73(97.26) | 0.14(0.13) | 21(22) |
| | RF + Stepwise (32 coefficients) | 99.91(99.92) | 0.02(0.02) | 21(22) |
| | Régression linéaire + Stepwise (16 coefficients) | 99.64(99.8) | 0.05(0.03) | 11(12) |
| | CART (16 coefficients) | 97.12(97.12) | 0.13(0.13) | 11(12) |
| | RF + Stepwise (16 coefficients) | 99.89(99.93) | 0.03(0.02) | 11(12) |
| | Régression linéaire + Stepwise (8 coefficients) | 98.32(99.62) | 0.08(0.05) | 7(8) |
| | CART + Stepwise (8 coefficients) | 97.31(96.7) | 0.12(0.14) | 7(8) |
| | RF + Stepwise (8 coefficients) | 99.66(99.84) | 0.04(0.03) | 7(8) |
| | Régression linéaire + Stepwise (4 coefficients) | 98.86(99.59) | 0.08(0.05) | 4(5) |
| | CART (4 coefficients) | 97.3(96.54) | 0.12(0.14) | 4(5) |
| RF + Stepwise (4 coefficients) | 99.35(99.74) | 0.06(0.04) | 4(5) | |
| Splines cubiques | Régression + Stepwise 30 polynômes | 99.53(99.77) | 0.05(0.04) | 29(30) |
| | CART + Stepwise 30 polynômes | 98.99(99.67) | 0.08(0.04) | 17(18) |
| | RF + Stepwise 30 polynômes | 94.9(99.63) | 0.17(0.05) | 7(8) |
| | Régression + Stepwise 10 polynômes | 78.42(99.27) | 0.35(0.06) | 5(6) |
| | CART + Stepwise 5 polynômes | 76.77(97.08) | 0.36(0.13) | 29(30) |
| | RF + Stepwise 5 polynômes | 80.62(97.18) | 0.33(0.13) | 17(18) |
| | Régression + Stepwise 20 polynômes | 93.92(97.26) | 0.19(0.12) | 7(8) |
| | CART + Stepwise 20 polynômes | 94.43(97.23) | 0.18(0.12) | 5(6) |
| | RF + Stepwise 20 polynômes | 85.97(99.42) | 0.28(0.06) | 29(30) |
| | Régression + Stepwise 10 polynômes | 95.28(99.46) | 0.16(0.05) | 17(18) |
| | CART + Stepwise 10 polynômes | 97.77(99.38) | 0.11(0.06) | 7(8) |
| | RF + Stepwise 10 polynômes | 95.73(99.5) | 0.15(0.05) | 5(6) |
| | Régression linéaire + Stepwise 5 B-spline | 98.68(99.54) | 0.09(0.05) | 4(5) |
| | CART + Stepwise 5 B-spline | 95.77(96.39) | 0.15(0.14) | 4(5) |
| | RF + Stepwise 5 B-spline | 99.18(99.58) | 0.07(0.05) | 4(5) |
| | Régression linéaire + Stepwise 10 B-spline | 99.55(99.72) | 0.05(0.04) | 8(9) |
| | CART + Stepwise 10 B-spline | 96.54(97.14) | 0.14(0.13) | 8(9) |
| | RF + Stepwise 10 B-spline | 99.61(99.87) | 0.05(0.03) | 8(9) |
| | Régression linéaire + Stepwise 20 B-spline | 99.89(99.91) | 0.02(0.02) | 14(15) |
| | CART + Stepwise 20 B-spline | 96.85(96.81) | 0.13(0.13) | 14(15) |
| RF + Stepwise 20 B-spline | 99.79(99.82) | 0.03(0.03) | 14(15) | |
| Régression linéaire + Stepwise 30 B-spline | 99.92(99.92) | 0.02(0.02) | 23(24) | |
| CART + Stepwise 30 B-spline | 96.55(96.65) | 0.14(0.14) | 23(24) | |
| RF + Stepwise 30 B-spline | 99.83(99.88) | 0.03(0.03) | 23(24) | |

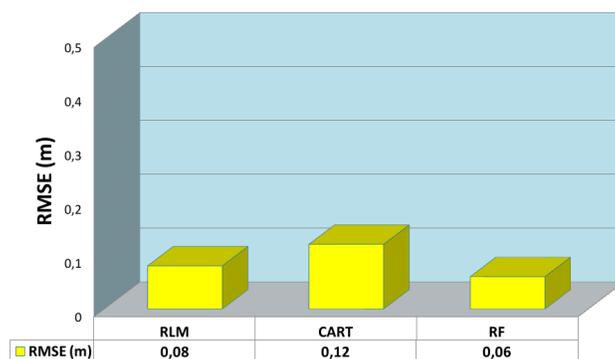
TABLE 12.3 – Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR simulées avec une eau claire.

A partir du tableau 12.3, on choisit les meilleurs modèles statistiques (Régression linéaire multiple, CART et Random Forest) parmi trois méthodes (l'Analyse en Composantes Principales, l'analyse en ondelettes et les B-Splines cubiques) qui contiennent moins de 5 variables explicatives (Les splines cubiques contiennent plus de 4 variables explicatives). J'expose sous forme de diagrammes à barres les meilleurs modèles statistiques⁶ pour chacune des méthodes utilisées.

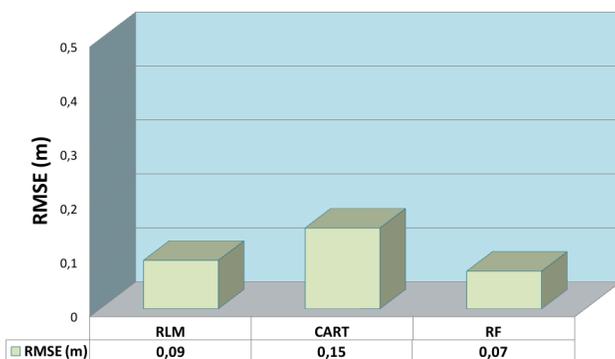
Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour l'ACP



Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour l'analyse en ondelettes



Les meilleurs résultats avec RLM, CART et RF (variables explicatives < 5 variables) pour les splines cubiques



6. Régression linéaire multiple, CART et Random Forest

Si on compare le RMSE des meilleurs modèles statistiques pour chacune des méthodes, on constate :

- qu'on obtient des résultats proches pour Random Forest et la régression linéaire multiple
- et que CART est encore une fois le modèle statistique le moins précis.

Notons également que les résultats des inversions des formes d'onde simulées sont plus précis que les résultats des inversions des formes d'onde réelles.

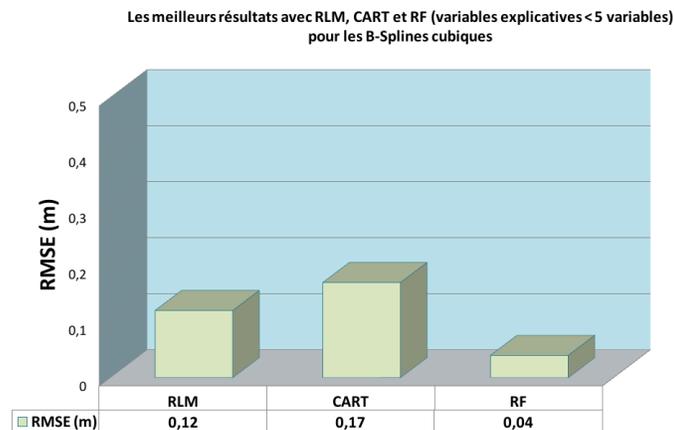
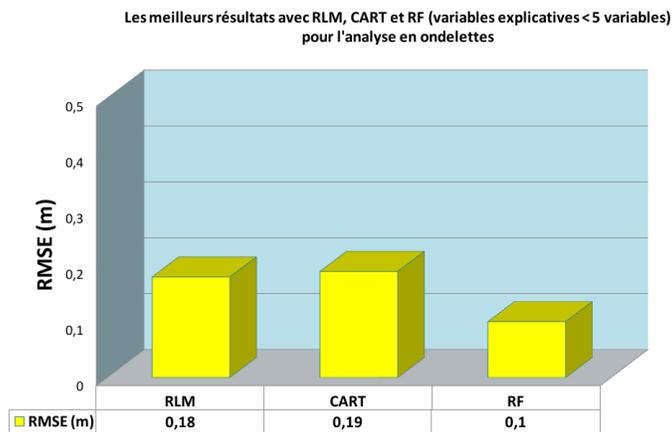
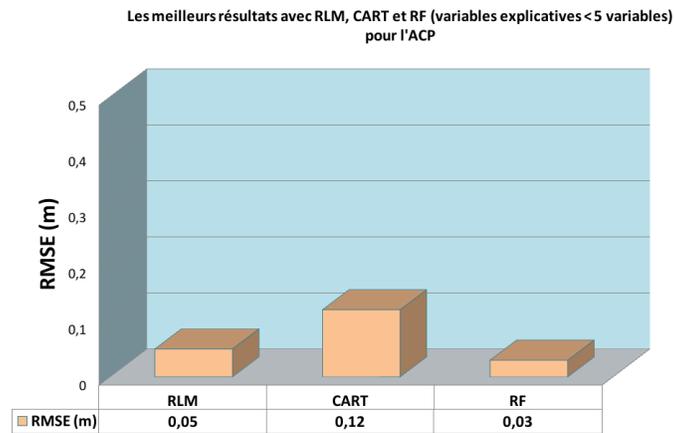
2.2 Résultats pour les données LiDAR avec une eau turbide

La figure 12.1 fournit le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) pour les données LiDAR simulées avec une eau turbide.

| | | Bathymétrie | | |
|--|--|-----------------------|-------------|---------------------|
| | | R ² (en %) | RMSE (en m) | Nombre de variables |
| ACP | Regression linéaire PCl | 41,9 | 0,57 | 1 |
| | Regression linéaire PCl-PC2 | 53,99 | 0,51 | 2 |
| | Regression linéaire PCl-PC2-PC3 | 59,48 | 0,48 | 3 |
| | Regression linéaire PCl-west | 59,47 | 0,05 | 2 |
| | Regression linéaire PCl-PC2-west | 59,47 | 0,05 | 3 |
| | Regression linéaire PCl-PC2-PC3-west | 59,52 | 0,05 | 4 |
| | CART PCl | 85,39 | 0,29 | 1 |
| | CART PCl-PC2 | 96,64 | 0,14 | 2 |
| | CART PCl-PC2-PC3 | 96,74 | 0,14 | 3 |
| | CART PCl-west | 96,63 | 0,14 | 2 |
| | CART PCl-PC2-west | 97,28 | 0,12 | 3 |
| | CART PCl-PC2-PC3-west | 96,46 | 0,14 | 4 |
| | RF PCl | 90,79 | 0,23 | 1 |
| | RF PCl-PC2 | 99,38 | 0,06 | 2 |
| | RF PCl-PC2-PC3 | 99,44 | 0,06 | 3 |
| | RF PCl-west | 99,77 | 0,04 | 2 |
| | RF PCl-PC2-west | 99,81 | 0,03 | 3 |
| | RF PCl-PC2-PC3-west | 99,75 | 0,04 | 4 |
| | Regression linéaire + Stepwise | 98,13 (99,74) | 0,10 (0,04) | 26 (27) |
| | CART + stepwise | 96,64 (96,61) | 0,14 (0,13) | 26 (27) |
| RF + stepwise | 99,78 (99,93) | 0,04 (0,02) | 26 (27) | |
| Ondelettes | Regression linéaire + Stepwise (256 coefficients) | | | |
| | CART + Step wise (256 coefficients) | | | |
| | RF + Stepwise (256 coefficients) | | | |
| | Regression linéaire + Stepwise (128 coefficients) | | | |
| | CART + Step wise (128 coefficients) | | | |
| | RF (128 coefficients) | | | |
| | Regression linéaire + Stepwise (64 coefficients) | 98,93 (99,44) | 0,08 (0,04) | 40 (41) |
| | CART + Step wise (64 coefficients) | 96,5 (96,27) | 0,14 (0,14) | 40 (41) |
| | RF + Stepwise (64 coefficients) | 99,71 (99,84) | 0,04 (0,03) | 40 (41) |
| | Regression linéaire + Stepwise (32 coefficients) | 99,5 (99,81) | 0,05 (0,03) | 23 (24) |
| | CART + Step wise (32 coefficients) | 96,03 (96,19) | 0,15 (0,15) | 23 (24) |
| | RF + Stepwise (32 coefficients) | 99,63 (99,8) | 0,05 (0,03) | 23 (24) |
| | Regression linéaire + Stepwise (16 coefficients) | 99,03 (99,77) | 0,07 (0,04) | 13 (14) |
| | CART (16 coefficients) | 96,48 (97,01) | 0,14 (0,13) | 13 (14) |
| | RF + Stepwise (16 coefficients) | 99,67 (99,99) | 0,05 (0,04) | 13 (14) |
| | Regression linéaire + Stepwise (8 coefficients) | 97,03 (99,31) | 0,15 (0,05) | 7 (8) |
| | CART + Stepwise (8 coefficients) | 95,53 (96,35) | 0,18 (0,14) | 7 (8) |
| | RF + Stepwise (8 coefficients) | 99,83 (99,95) | 0,03 (0,03) | 7 (8) |
| | Regression linéaire + Stepwise (4 coefficients) | 94,23 (99,35) | 0,18 (0,08) | 4 (5) |
| | CART (4 coefficients) | 93,27 (96,14) | 0,19 (0,14) | 4 (5) |
| RF + Stepwise (4 coefficients) | 99,5 (99,49) | 0,10 (0,05) | 4 (5) | |
| Splines cubiques | Regression +Stepwise 30 polynomes | 98,8 (99,71) | 0,08 (0,04) | 29 (30) |
| | Regression +Stepwise 20 polynomes | 97,59 (99,62) | 0,12 (0,05) | 21 (22) |
| | Regression +Stepwise 10 polynomes | 95,41 (99,46) | 0,21 (0,08) | 11 (12) |
| | Regression +Stepwise 5 polynomes | 85,59 (99,42) | 0,28 (0,06) | 5 (6) |
| | CART+Stepwise 30 polynomes | 87,57 (99,36) | 0,26 (0,14) | 29 (30) |
| | CART+Stepwise 20 polynomes | 88,43 (96,15) | 0,26 (0,15) | 21 (22) |
| | CART+Stepwise 10 polynomes | 88,40 (96,48) | 0,25 (0,14) | 11 (12) |
| | CART+Stepwise 5 polynomes | 88,53 (96,81) | 0,25 (0,13) | 5 (6) |
| | RF +Stepwise 30 polynomes | 97,84 (99,83) | 0,11 (0,05) | 29 (30) |
| | RF +Stepwise 20 polynomes | 97,91 (99,48) | 0,11 (0,05) | 21 (22) |
| | RF +Stepwise 10 polynomes | 98,03 (99,48) | 0,10 (0,06) | 11 (12) |
| | RF +Stepwise 5 polynomes | 97,02 (99,89) | 0,13 (0,05) | 5 (6) |
| | Regression linéaire+Stepwise 5 B-spline | 97,61 (99,48) | 0,12 (0,05) | 4 (5) |
| | CART+Stepwise 5 B-spline | 95,04 (96,25) | 0,17 (0,13) | 4 (5) |
| | RF +Stepwise 5 B-spline | 99,63 (99,82) | 0,04 (0,03) | 4 (5) |
| | Regression linéaire+Stepwise 10 B-spline | 99,94 (99,71) | 0,03 (0,04) | 9 (10) |
| | CART+Stepwise 10 B-spline | 94,03 (96,81) | 0,18 (0,13) | 9 (10) |
| | RF +Stepwise 10 B-spline | 99,43 (99,77) | 0,06 (0,04) | 9 (10) |
| | Regression linéaire+Stepwise 20 B-spline | 99,52 (99,81) | 0,05 (0,03) | 14 (15) |
| | CART+Stepwise 20 B-spline | 95,43 (96,8) | 0,16 (0,13) | 14 (15) |
| RF +Stepwise 20 B-spline | 99,74 (99,85) | 0,04 (0,03) | 14 (15) | |
| Regression linéaire+Stepwise 30 B-spline | 99,53 (99,83) | 0,05 (0,05) | 23 (24) | |
| CART+Stepwise 30 B-spline | 94,15 (95,02) | 0,18 (0,17) | 23 (24) | |
| RF +Stepwise 30 B-spline | 99,63 (99,68) | 0,05 (0,04) | 23 (24) | |

FIGURE 12.1 – Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR simulées avec une eau turbide.

A partir du tableau 12.1, on choisit les meilleurs modèles statistiques (Régression linéaire multiple, CART et Random Forest) parmi trois méthodes (l'Analyse en Composantes Principales, l'analyse en ondelettes et les B-Splines cubiques) qui contiennent moins de 5 variables explicatives (Les splines cubiques contiennent plus de 4 variables explicatives). J'expose sous forme de diagrammes à barres les meilleurs modèles statistiques⁷ pour chacune des méthodes utilisées.



7. Régression linéaire multiple, CART et Random Forest

Si on compare le RMSE des meilleurs modèles statistiques pour chacune des trois méthodes, on constate :

- qu'on obtient des résultats proches pour Random Forest et la régression linéaire multiple
- et que CART est encore une fois le modèle statistique le moins précis.

3 Résultats pour les données LiDAR altimétriques du Brésil

La figure 12.4 fournit le coefficient de détermination R^2 (%), le RMSE (m) et le nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) pour les données LiDAR altimétriques (forêt brésil).

| | | Hauteur | | |
|---|--|----------------|-------------|---------------------|
| | | Brésil [0-30m] | | |
| | | R^2 (en %) | RMSE (en m) | Nombre de variables |
| ACP | Régression linéaire PCL | 12.91 | 0.00 | 1 |
| | Régression linéaire PCL-PC2 | 14.41 | 0.04 | 2 |
| | Régression linéaire PCL-PC2-PC3 | 17.79 | 5.93 | 3 |
| | Régression linéaire PCL-west | 62.57 | 3.99 | 2 |
| | Régression linéaire PCL-PC2-west | 64.06 | 3.91 | 3 |
| | Régression linéaire PCL-PC2-PC3-west | 65.03 | 3.86 | 4 |
| | CART PCL | 13.88 | 6.18 | 1 |
| | CART PCL-PC2 | 14.11 | 6.32 | 2 |
| | CART PCL-PC2-PC3 | 16.5 | 6.42 | 3 |
| | CART PCL-west | 71.76 | 3.48 | 2 |
| | CART PCL-PC2-west | 72.82 | 3.41 | 3 |
| | CART PCL-PC2-PC3-west | 70.95 | 3.56 | 4 |
| | RF PCL | 7.4 | 6.77 | 1 |
| | RF PCL-PC2 | 17.86 | 6.05 | 2 |
| | RF PCL-PC2-PC3 | 41.76 | 4.98 | 3 |
| | RF PCL-west | 75.07 | 3.26 | 2 |
| | RF PCL-PC2-west | 81 | 2.94 | 3 |
| | RF PCL-PC2-PC3-west | 81.47 | 2.97 | 4 |
| | Régression linéaire + 5te pwise | 33.62(72.95) | 5.37(3.4) | 17(18) |
| | CART + stepwise | 61.01(76.15) | 4.14(3.21) | 17(18) |
| RF + stepwise | 78.03(86.09) | 3.21(2.52) | 17(18) | |
| Ondelettes | Régression linéaire + Stepwise (256 coefficients) | | | |
| | RF + Stepwise (256 coefficients) | | | |
| | Régression linéaire + Stepwise (128 coefficients) | 51.42(72.55) | 4.64(3.46) | 48(49) |
| | CART + Stepwise (128 coefficients) | 58.51(76.41) | 4.23(3.18) | 48(49) |
| | RF (128 coefficients) | 81.22(87.08) | 2.97(2.47) | 48(49) |
| | Régression linéaire + Stepwise (64 coefficients) | 44.06(68.29) | 4.93(3.69) | 19(20) |
| | CART + Stepwise (64 coefficients) | 63.43(72.01) | 4.06(3.40) | 19(20) |
| | RF + Stepwise (64 coefficients) | 73.44(84.11) | 3.52(2.75) | 19(20) |
| | Régression linéaire + Stepwise (32 coefficients) | 37.46(65.42) | 5.2(3.84) | 9(10) |
| | CART + Stepwise (32 coefficients) | 68.54(76.39) | 3.67(3.2) | 9(10) |
| | RF + Stepwise (32 coefficients) | 76.02(82.97) | 3.25(2.74) | 9(10) |
| | Régression linéaire + Stepwise (16 coefficients) | 39.46(61.18) | 5.09(4.09) | 7(8) |
| | CART (16 coefficients) | 67.01(73.91) | 3.9(3.37) | 7(8) |
| | RF + Stepwise (16 coefficients) | 82.68(88.46) | 2.69(2.54) | 7(8) |
| | Régression linéaire + Stepwise (8 coefficients) | 34.73(64.25) | 5.24(3.9) | 4(5) |
| | CART + Stepwise (8 coefficients) | 51.12(61.81) | 4.68(4.08) | 4(5) |
| | RF + Stepwise (8 coefficients) | 72.05(76.7) | 3.5(3.21) | 4(5) |
| | Régression linéaire + Stepwise (4 coefficients) | 35.8(58.74) | 5.23(4.19) | 2(3) |
| | CART (4 coefficients) | 54.78(66.17) | 4.46(3.8) | 2(3) |
| | RF + Stepwise (4 coefficients) | 67.08(74.18) | 3.75(3.32) | 2(3) |
| Régression +Stepwise 30 polynomes | 39.68(72.81) | 5.33(3.45) | 30(31) | |
| Régression +Stepwise 20 polynomes | 34.24(76.35) | 5.42(3.19) | 18(19) | |
| Régression +Stepwise 10 polynomes | 25.19(71.13) | 5.68(3.52) | 8(9) | |
| Régression +Stepwise 5 polynomes | 14.61(60.33) | 6.05(4.11) | 4(5) | |
| CART +Stepwise 30 polynomes | 48.11(66.34) | 4.82(3.84) | 30(31) | |
| CART +5stepwise 20 polynomes | 46.44(64.76) | 4.96(3.94) | 18(19) | |
| CART +5stepwise 10 polynomes | 38.41(61.04) | 5.87(4.14) | 8(9) | |
| CART +5stepwise 5 polynomes | 18.18(70.43) | 6.35(3.56) | 4(5) | |
| RF +5stepwise 30 polynomes | 68.62(78.5) | 3.83(3.16) | 30(31) | |
| RF +5stepwise 20 polynomes | 59.85(76.4) | 4.19(3.26) | 18(19) | |
| RF +5stepwise 10 polynomes | 41.92(77.14) | 4.98(3.22) | 8(9) | |
| RF +5stepwise 5 polynomes | 21.36(72.11) | 5.91(3.47) | 4(5) | |
| Régression linéaire+5te pwise 5 B-spline | 39.44(63.24) | 5.08(3.96) | 2(3) | |
| CART +5stepwise 5 B-spline | 70.84(79.36) | 3.54(2.98) | 2(3) | |
| RF +5stepwise 5 B-spline | 73.92(87.39) | 3.34(2.35) | 2(3) | |
| Régression linéaire+5te pwise 10 B-spline | 41.63(62.79) | 4.99(3.99) | 4(5) | |
| CART +5stepwise 10 B-spline | 66.33(76.57) | 3.83(3.19) | 4(5) | |
| RF +5stepwise 10 B-spline | 68.86(84.25) | 3.73(2.75) | 4(5) | |
| Régression linéaire+5te pwise 20 B-spline | 57.69(67.66) | 5.23(3.77) | 12(13) | |
| CART +5stepwise 20 B-spline | 65.15(73.14) | 3.94(3.43) | 12(13) | |
| RF +5stepwise 20 B-spline | 80.08(83.43) | 3.03(2.58) | 12(13) | |
| Régression linéaire+5te pwise 30 B-spline | 45.54(72.43) | 4.84(3.46) | 18(17) | |
| CART +5stepwise 30 B-spline | 57.87(70.37) | 4.33(3.6) | 18(17) | |
| Splines cubiques | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| B-splines | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |
| | | | | |

TABLE 12.4 – Coefficient de détermination R^2 (%), RMSE (m) et nombre de variables explicatives des différents modèles statistiques (Régression linéaire multiple, CART et Random Forest) avec les différentes méthodes utilisées (Analyse en Composantes Principales, analyse en ondelettes, splines cubiques et B-Splines cubiques) Les résultats entre parenthèses intègrent aux modèles statistiques la variable explicative W_{ext} (partie utile de la forme d'onde) - Données LiDAR altimétriques (forêt brésil).

A partir du tableau 12.4, on choisit les meilleurs modèles statistiques (Régression linéaire multiple, CART et Random Forest) parmi les quatre méthodes (l'Analyse en Composantes Principales, l'analyse en ondelettes, les splines cubiques et les B-Splines) qui contiennent moins de 5 variables explicatives. J'expose

sous forme de diagrammes à barres les meilleurs modèles statistiques⁸ (construits avec moins de 5 variables explicatives) pour chacune des méthodes utilisées.

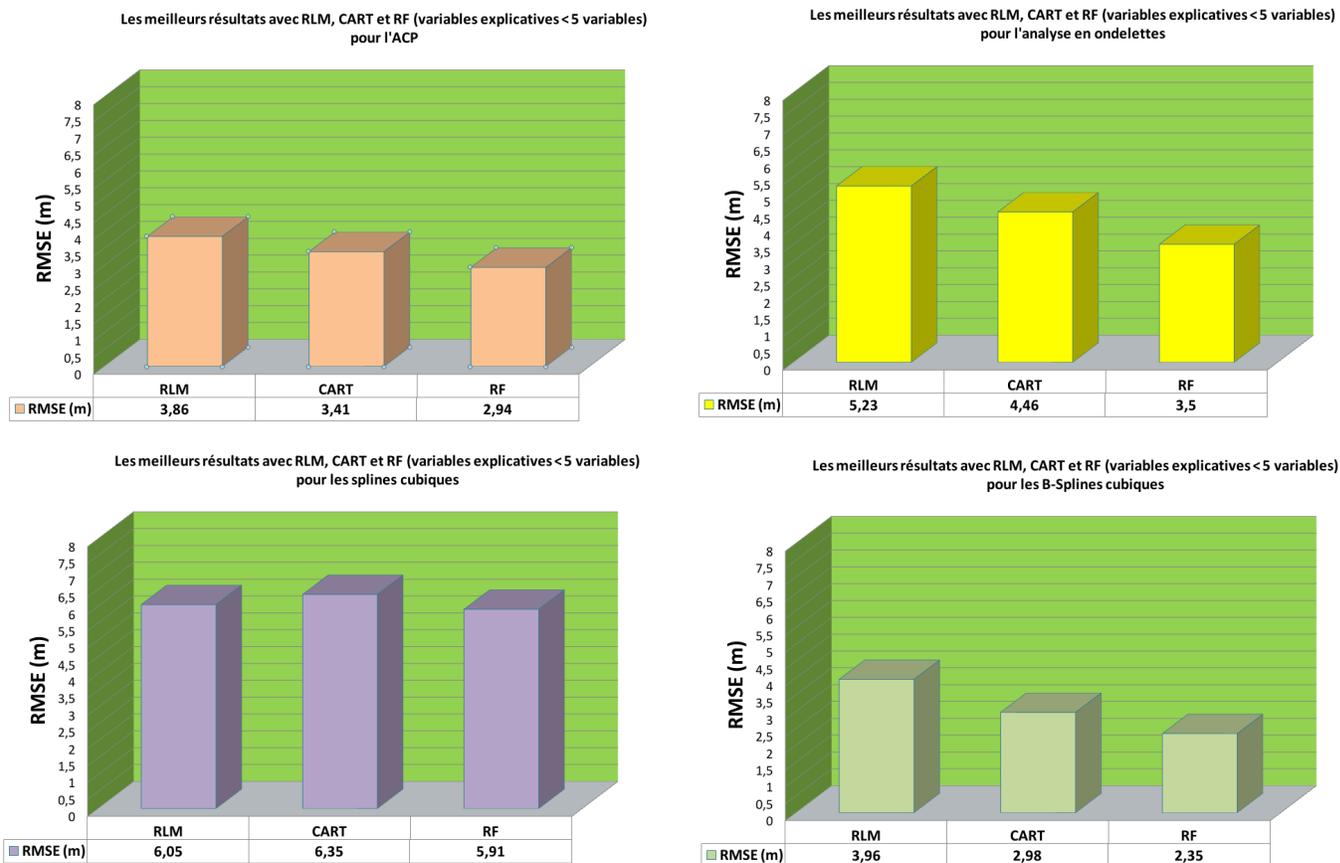


TABLE 12.5 – Forêt Brésil - Les meilleurs résultats (RMSE (m)) de la figure 12.1 avec la régression linéaire multiple, CART et Random Forest avec un nombre de variables explicatives inférieur à 5 en utilisant l'Analyse en Composantes Principales, l'analyse en ondelette, les splines cubiques et les B-Splines cubiques.

Si on compare les résultats des meilleurs modèles statistiques pour chaque méthode, on observe qu'on obtient de meilleurs résultats avec Random Forest (Régression non linéaire). On arrive à obtenir une erreur quadratique moyenne (RMSE) de 2.35 m pour les B-Splines cubiques. On remarque aussi que l'inversion des formes d'onde avec la méthode des splines cubiques est la moins précise. Et on peut également noter une amélioration des résultats (Tableau 12.4) en ajoutant la variable explicative W_{ext} (partie utile d'une forme d'onde).

8. Régression linéaire multiple, CART et Random Forest

Chapitre 13

Conclusion

Au cours de ce stage, j'ai développé de nouvelles méthodes d'estimation (d'inversion) de la bathymétrie et de la hauteur d'arbres à partir de formes d'onde LiDAR. Ces nouvelles méthodes statistiques, utilisant un apprentissage, tombent dans le domaine dit de **l'analyse fonctionnelle**. Ces méthodes **réduisant le signal en un nombre réduit de paramètres** qui décrivent ce signal et une régression statistique (modèle d'inversion) est ensuite effectuée entre **ces paramètres et la variable d'intérêt** (bathymétrie ou hauteur de canopée).

Quatre méthodes de réduction de la dimension ont été testées : **l'Analyse en Composantes Principales, la décomposition en ondelettes, l'approximation par splines cubique et l'approximation par B-Splines**.

Trois méthodes de régression ont été testées : **la régression multiple, CART et la méthode de forêt aléatoire (random forest)**. Les résultats de ces méthodes montrent que, pour des formes d'onde simulées, parfaitement contrôlées, on peut estimer la bathymétrie avec des précisions allant jusqu'à 3 cm pour des gammes de profondeur comprises entre 0 et 3 mètres. Les résultats de ces méthodes sur des formes d'onde simulées pour une eau claire et turbide sont équivalents.

Les résultats des modèles d'inversion des formes d'onde bathymétriques réelles pour le Golfe du Morbihan avec deux types de capteurs (APD et PMT) sont également très proches. L'écart de précision entre les formes d'onde réelles et simulées s'explique par le fait que les paramètres du milieu sont fixés pour le cas des formes d'onde LiDAR simulées. En ce qui concerne l'estimation des hauteurs d'arbres, les résultats des modèles d'inversion des formes d'onde LiDAR atlimétriques (réelles) obtenues donnent une précision allant jusqu'à 2.35 mètres.

Les modèles d'inversion utilisant **les forêts aléatoires donne de meilleurs résultats** que pour la régression linéaire. CART reste le modèle statistique le moins précis. Si on compare les différentes méthodes, on voit que **les B-Splines donnent les meilleurs résultats**. L'Analyse en Composantes Principales et l'analyse en ondelettes restent également un bon choix pour l'inversion des formes d'onde LiDAR. L'avantage de ces méthodes est qu'elles proposent des modèles statistiques avec peu de variables explicatives.

Ce stage a été très intéressant du fait des différentes disciplines associées qui m'ont permis de progresser dans plusieurs domaines. Tout d'abord dans le domaine de l'informatique, avec la prise en main d'un nouveau langage de programmation interprété dédié à l'analyse statistique : **R** avec l'utilisation de nombreux packages. J'ai également appris beaucoup de méthodes en analyse statistique des données et traitement du signal. Enfin, j'ai été initié à la télédétection active (laser), ce qui m'a permis de découvrir un domaine qui m'était jusqu'alors inconnu.

Cinquième partie

ANNEXES

Annexe A

Le logiciel R



est **distribué gratuitement** à partir du site du **CRAN**. R est un système d'**analyse statistique** créé par Ross Ihaka et Robert Gentleman **distribué librement sous les termes de la GNU** (General Public Licence); son développement et sa distribution sont assurés par plusieurs statisticiens rassemblés dans le R Development Core Team. Le **langage R** est un langage de programmation et un environnement mathématique utilisés pour l'analyse statistique. R dispose de **nombreuses fonctions graphiques**. Les sites miroirs proposent des centaines de bibliothèques sous **licence libre**, et leur nombre croît régulièrement. R est basé sur le langage S qui a été développé par John Chambers des laboratoires Bell. R est un langage avant tout procédural à destination de statisticiens : il possède par conséquent une bibliothèque de modèles déjà implémentés extrêmement intéressante. **L'analyse des données** n'est pas oubliée avec l'implémentation de nombreuses **méthodes de classification**. Le module graphique est quant à lui aussi très intéressant : il permet de faire un grand nombre de types de graphiques (histogrammes, camemberts,...).

Annexe B

Simulateur Wa-LID 2.0

```

1
2 # WaLID program for flat bottom
3 # J.S. Bailly , M. Montes, A. Bouhdaoui, N. Baghdadi, December 2013
4 # M. Mezian 2014
5 # AgroParisTech – UQAR – IRSTEA
6
7 ##### Repertoire de travail et packages #####
8
9
10
11 rm(list = ls())
12 setwd("/homeL/mezian/R/WaLiD_R_122013_JeanStephane_underconstruction")
13 require(gdata)
14
15 ##### Parametres du capteur #####
16
17 T0=3.5e-09
18 fs=1e+09
19 theta_nadir=0.01
20 H=500000
21 c=2.99705e+08
22 k=6
23 Zmax=55.15
24 E0<-5e-05
25 Ar=0.63
26 lambda=532
27 Delta.B<-1e-9
28 FOV<-9e-05
29 e<-1.6e-19
30 G_noise<-3
31 Id<-1e-10
32 R.lambda<-0.25
33 Mgauss<-4.6
34 G<-13.34
35 gamma=15e-3
36 F<-1
37 eta.e=0.8
38 eta.r=0.5
39 Fp<-H*gamma
40
41
42 B<-1/(2*T0)
43 Pe<-E0/T0
44 if (lambda >=315 & lambda <=400){EMP=5600*(T0^0.25)}
45 if (lambda >400 & lambda <=700){EMP=5e-3;}
46 if (lambda >700 & lambda <=1050){EMP=10^((lambda-700)/500)*5e-3}

```

```

47 if (lambda > 1050 & lambda <= 1400) {EMP=5e-2}
48 gamma.r=0.01
49 eta.lambda=0.07
50
51
52 ##### Parametres air #####
53
54 param.wp<-read.xls(paste(getwd(),"/water.xls",sep=""),sheet = 1, header = TRUE)
55
56
57
58
59 T2.atm<-NULL
60
61 if (is.null(T2.atm)) {T2.atm<-approx(param.wp[,"Lambda"],param.wp[,"Tatm"],xout=lambda)$y }
62
63 wind<-10
64 direc<-"along"
65 Mscint<-2
66 Is<-0.025
67
68
69 if (E0 > ( (EMP*pi*Fp^2) / (sqrt(T2.atm)*Mscint*Mgauss*4*(G^2)) ) ) {print("Attention : Vous
    devez choisir E0 de tel sorte qu'il respecte la contrainte de surete")}
70
71 ##### Parametres modelisation #####
72 BRDF="CT" # valeur 'CT' ou 'CM'
73
74 ##### Parametres eau #####
75 nw=1.333
76 epsilon=0.1
77 Z<-4
78 Beta=1.4e-03
79 M=0.014
80 g=1.14e-02
81 lambda.0=440
82 ks=0.84434855
83 theta.s=0.000194494
84 ay0=0.01
85 S=2
86 Cph=0.5
87 Rb=0.04
88 r=0.23032251
89
90 K=NULL
91 if (is.null(K)) {
92
93     if (lambda < 400 | lambda > 700) {stop("Attention la valeur de la longueur d'onde 'lambda' doit
        etre compris entre 400 et 700nm ")}
94
95
96     a.wp=approx(param.wp[,"Lambda"],param.wp[,"coeff.absorption"],xout=lambda)$y
97
98     c.phyto=approx(param.wp[,"Lambda"],param.wp[,"coeff.normalise.des.phytoplanctons"],xout=
        lambda)$y
99
100     a.phyto=(0.06*c.phyto*(Cph^0.602))
101
102     a.y=(ay0*exp(-M*(lambda-lambda.0)))
103
104     a.s=(0.042*S/3.327)*exp(-g*(lambda-lambda.0))
105
106     a.w=a.wp+a.phyto+a.s+a.y
107
108     b.wp=approx(param.wp[,"Lambda"],param.wp[,"coeff.diffusion"],xout=lambda)$y

```

```

109  b.s=(S/3.327)*(lambda/550)^{-1}
110  b.phyto=0.12*Cph^0.63*0.357/c.phyto
111  b.y=0
112
113  b.w=b.wp+b.phyto+b.y+b.s
114
115  c.w=a.w+b.w
116  w0=b.w/c.w
117  if(w0<=0 & w0 >=1){stop("Attention il faut que : 0<Wo<1 ")}
118
119  K=c.w*(0.19*(1-w0))^(w0/2)
120  if(K<0.1 |K>0.55){stop("Attention il faut que : 0.1=<K<=0.55 ")}
121
122 }
123
124
125 ##### Les angles #####
126
127 theta.b<-0
128
129 theta=theta_nadir-theta.s
130
131 theta.t<-asin(sin(theta)/nw)
132
133 #####Echelle de temps#####
134
135 ts<-(2*H)/(c*cos(theta))
136
137 ti<-ts-k*T0
138
139 t.lag<-1/(sefs)
140
141 serie1<-c(seq(ti,ts,by=t.lag),ts)
142
143 n.sample.w<-floor(((2*Zmax)/((c/nw)*cos(theta.t)))/t.lag)
144
145 tf<-ts+n.sample.w*t.lag
146
147 serie2<-seq(ts+t.lag,tf,by=t.lag)
148
149 serie<-c(serie1,serie2)
150
151 tb<-ts+(2*Z/(c/nw)*cos(theta.t))
152
153 tcs<-serie[serie>ts & serie < tb]
154
155 ##### Difference de temps #####
156
157 delta.t.s<-((2*H)/c)*(1/(cos((theta)+(gamma/2)))-1/(cos(theta)-(gamma/2)))
158
159 delta.t.b<-((2*Z)/((c/nw)*cos(asin(sin(theta+gamma/2))/nw))-
160 (2*Z)/((c/nw)*cos(asin(sin(theta-gamma/2))/nw)))
161
162 delta.t<-delta.t.s+delta.t.b
163
164 ##### Test de Guenther #####
165
166
167 if(epsilon<0.1 | epsilon >0.35){stop("!!-----Attention, epsilon doit verifier : 0.1<=
168 epsilon <=0.35 -----!!")}
169
170 ifelse(delta.t.s<=(2*T0),T1<-T0+epsilon*delta.t.s,T1<-T0*(1+((0.5*delta.t.s)/T0-0.4)))
171
172 ifelse(delta.t.b<=(2*T1),T2<-T1+epsilon*delta.t.b,T2<-T1*(1+((0.5*delta.t.b)/T1-0.4)))

```

```

172
173 ##### Ecartstypes #####
174
175 sigma<-T0/(2*sqrt(2*log(2)))
176
177 sigma1<-T1/(2*sqrt(2*log(2)))
178
179 sigma2<-T2/(2*sqrt(2*log(2)))
180
181
182 ##### Facteur de perte : Modele CM ou CT #####
183
184
185 if (BRDF=="CT"){
186
187   if (r<0 | r>0.5){stop("Attention, !!-----r doit varier entre 0 et
188   0.5-----!!")}
189
190   if (ks<0 | ks>1){stop("!!-----Attention, il faut que 0<=ks
191   <=1-----!!")}
192
193   D<-exp(-(tan(theta)/r)^2)/(r^2*cos(theta)^4)
194
195   O<-min(1,2*cos(theta)^2)
196
197   ifelse(theta==0,Fr<-(nw-1)^2 / (nw+1)^2,Fr<-0.5*((sin(theta-theta.t)/sin(theta+theta.t))
198   )^2 + (tan(theta-theta.t)/tan(theta+theta.t))^2)
199
200   Ls<-(1-ks)/pi + (ks*D*O*Fr)/(pi*cos(theta)^2)
201
202 }
203
204 if (BRDF=="CM") {
205
206   ifelse(wind<=17,delta.s<-214*(1-(wind-4.8)/12.2)^(2.06)+37,delta.s<-4.26*exp(-0.14*wind)
207   )
208
209   if (wind<=3){stop("!!----Attention la valeur de la vitesse du vent dans le modele de Cox-
210   Munk ne peut etre inferieur ou egale a 3-----!!")}
211
212   ifelse(direc=="along",sigmad<-0.00158*wind,sigmad<-0.003+0.00096*wind)
213
214   if (direc!="along" & direc!="across"){stop("!! —— Attention : Vous devez choisir la
215   direction du vent direc='along' ou direc='across (Pas d'autres choix possible)
216   -----!!")}
217
218   Nd<-exp(-theta^2)/(2*sigmad^2)
219
220   Ls<-delta.s*Nd*(cos(theta)^2)
221
222 }
223
224 if (BRDF!="CM" & BRDF!="CT"){stop("!! —— Attention : Vous devez choisir le modele Cook-
225   Torrance (BRDF='CT') ou Cox-Munk (BRDF='CM') -----!!")}
226
227 ##### Choisir Type de fonction #####
228
229 distribution_gaussienne=function(t,tx,Ecart_type){
230   distribution_gaussienne=(2*Pe/sqrt(2*pi))*exp(-(t-tx)^2/(2*Ecart_type^2))
231   return (distribution_gaussienne)
232 }

```

```

229 }
230 ##### Forme d'onde emise par le capteur #####
231
232 wfe=distribution_gaussienne(serie ,ts ,sigma)
233
234 ##### Forme de l'onde retourne par la surface #####
235
236 wfe.s=distribution_gaussienne(serie ,ts ,sigma1)
237
238 Ps<-Pe*(T2.atm*Ar*eta.e*eta.r*Ls*cos(theta)^2)/(pi*H^2)
239
240 P_s<-wfe.s*Ps
241
242 ##### Forme de l'onde retourne par la colonne d'eau #####
243
244 zb<-(tcs-ts)*(c/nw)*cos(theta.t)/2
245
246
247 wfe.c=distribution_gaussienne(serie ,ts ,sigma1)
248
249 if(F<0|F>1){stop("!!-----Attention , il faut que 0<=F<=1-----!!")}
250
251 Pc<-((Pe*T2.atm*Ar*eta.e*eta.r*F*(1-Ls)^2*Beta*exp(-(2*K*zb)/cos(theta.t)))/(((nw*H)+zb)/cos(theta))^2)
252
253
254 temp<-rep(0 , length(serie))
255 temp[1:length(Pc)]<-Pc
256 Pc<-temp
257
258 P_c<-convolve(wfe.c , rev(Pc) , type="open")
259
260 P_c=P_c[1:length(serie)]
261
262 ##### Forme de l'onde retourne par le fond de l'eau #####
263
264 wfe.b=distribution_gaussienne(serie ,tb ,sigma2)
265 Pb<-Pe*T2.atm*Ar*eta.e*eta.r*F*Rb*(1-Ls)^2*exp(-2*K*Z/cos(theta.t)) / (pi*((nw*H+Z)/cos(theta))^2)
266
267
268 P_b<-wfe.b*Pb
269
270 ##### Bruit radiation solaire #####
271
272
273 Pbg=Is*sqrt(T2.atm)*Ar*(1-(gamma.r^2))*(pi*(FOV^2)/4)*Delta.B*eta.r*eta.lambda
274
275 P_bg<-rep(Pbg , length(serie))*rnorm(length(serie) , 0 , 1)
276
277
278
279 ##### Bruit capteur #####
280
281 PSNR=100;
282
283 sigmaN=max(P_s+P_c+P_b)/PSNR
284
285 sigmaN<-sigmaN[!is.na(sigmaN)]
286
287 P_N<-rnorm(length(serie) , 0 , sigmaN)
288
289
290 ##### Forme del'onde total #####
291

```

```

292 P_tot<-Pext+P_N
293
294
295
296 ##### Affichage #####
297
298
299
300 tmax<-100
301 y.limit=c(0, max(P_tot))
302 mat<-matrix(c(1:5, c(6,6,rep(7,3))), byrow=FALSE, ncol=2, nrow=5)
303 matnf<-layout(mat)
304 par(mar=c(4,4,4,4))
305 xlab="Time (t) ns "
306 ylab="Power (w)"
307
308 plot(P_s, xlim=c(0,tmax), ylim=c(0,max(P_s)), type='l',cex.main=0.9,cex.lab = 1,main="Water
  surface ", xlab=xlab,ylab=ylab,col="gray")
309 plot(P_c, xlim=c(0,tmax), ylim=c(0,max(P_c)), type='l',cex.main=0.9,cex.lab = 1,main="Water
  column ", xlab=xlab,ylab=ylab,col="blue")
310 plot(P_b, xlim=c(0,tmax), ylim=c(0,max(P_b)), type='l',cex.main=0.9,cex.lab = 1,main="
  Waterbottom ", xlab=xlab,ylab=ylab,col="brown")
311 plot(P_bg, xlim=c(0,tmax), ylim=c(min(P_bg),max(P_bg)), cex.main=0.9,cex.lab = 1,type='l',
  main="Solar noise", xlab=xlab,ylab=ylab,col="yellow")
312 plot(P_N, xlim=c(0,tmax), ylim=c(min(P_N),max(P_N)),cex.main=0.9,cex.lab = 1, type='l',main=
  "Detector noise ", xlab=xlab,ylab=ylab,col="red")
313 plot(wfe, xlim=c(0,tmax), ylim=c(0,max(wfe)*(1.05)),cex.main=0.9,cex.lab = 1, type='l',main=
  "Emmited waveform ", xlab=xlab,ylab=ylab,col="green")
314
315 plot(P_tot, xlim=c(0,90), ylim=c(0,3e-06), type='l',main="Back-scattered waveform ",cex.main
  =0.9,cex.lab = 1,xlab=xlab,ylab=ylab,col="red")
316
317
318
319 ##### Sauvegarde dans un fichier #####
320
321
322
323 write.table(P_tot,file ="Ptot.csv", append = TRUE, quote = FALSE, sep= ";", na="NA",row.
  names=FALSE,col.names=FALSE)

```

Annexe C

Reconstruire un signal à l'aide des ondelettes de Haar

```

1
2
3 setwd("/homeL/mezian/R/WaLiD_R_122013_JeanStephane_underconstruction/lidar_foret_bresil/
4 formes_ondes_brezil/raw_waveforme/")
5 donnees_morbihan1<-read.csv(file="toutGAPDechantillonne.csv",sep=" ",dec=".",header=F)
6
7 Z=donnees_morbihan1[189,seq(3,130,2)]
8
9 library(waveslim)
10 Haar_approximation_father_by_level=function(J,Z){
11   e=0
12   b=1
13   J=J
14   t0=0.5
15   a=dwt(Z,'haar',(log2(length(Z))/log2(2))-1)
16   t=seq(1,length(Z),1)
17   x=rep(0,length(t))
18   while(e<length(Z)){
19
20     indice1=which(t>=t0 & t <(2^J)/2+t0)
21     indice2=which(t>=(2^J)/2+t0 & t <=2^J+t0)
22
23     x[indice1]=-(2^(-J/2)) *a[[J]][b]
24     x[indice2]=(2^(-J/2)) *a[[J]][b]
25
26     t0=t0+2^J
27     e=e+2^J
28     b=b+1
29   }
30   return(x)
31 }
32 }
33
34 Haar_approximation_father_by_level_sans_coeff=function(J,Z){
35   e=0
36   b=1
37   J=J
38   t0=0.5
39   a=dwt(Z,'haar',(log2(length(Z))/log2(2))-1)
40   t=seq(1,length(Z),1)
41   x=rep(0,length(t))
42   while(e<length(Z)){

```

ANNEXE C. RECONSTRUIRE UN SIGNAL À L'AIDE DES ONDELETTES DE HAAR

```

43
44     indice1=which(t>=t0 & t <(2^J)/2+t0)
45     indice2=which(t>=(2^J)/2+t0 & t <=2^J+t0)
46
47     x[indice1]=-(2^(-J/2))
48     x[indice2]=(2^(-J/2))
49
50     t0=t0+2^J
51     e=e+2^J
52     b=b+1
53 }
54 return(x)
55
56 }
57
58
59
60 Haar_approximation_mother_by_level=function(J,Z){
61     e=0
62     b=1
63     J=J
64     t0=1
65     a=dwt(Z,'haar',(log2(length(Z))/log2(2))-1)
66     t=seq(1,length(Z),1)
67     x=rep(0,length(t))
68     while(e<length(Z)){
69
70         indice1=which(t>=t0 & t <(2^J)/2+t0)
71         indice2=which(t>=(2^J)/2+t0 & t <=2^J+t0)
72         x[indice1]=(2^(-J/2))*a[[length(a)]] [b]
73         x[indice2]=(2^(-J/2))*a[[length(a)]] [b]
74         t0=t0+2^J
75         e=e+2^J
76         b=b+1
77     }
78
79     return(x)
80
81 }
82
83
84 Haar_approximation_mother_by_level_sans_coeff=function(J,Z){
85     e=0
86     b=1
87     J=J
88     t0=1
89     a=dwt(Z,'haar',(log2(length(Z))/log2(2))-1)
90     t=seq(1,length(Z),1)
91     x=rep(0,length(t))
92     while(e<length(Z)){
93
94         indice1=which(t>=t0 & t <(2^J)/2+t0)
95         indice2=which(t>=(2^J)/2+t0 & t <=2^J+t0)
96         x[indice1]=(2^(-J/2))
97         x[indice2]=(2^(-J/2))
98         t0=t0+2^J
99         e=e+2^J
100        b=b+1
101    }
102
103    return(x)
104
105 }
106
107

```

ANNEXE C. RECONSTRUIRE UN SIGNAL À L'AIDE DES ONDELETTES DE HAAR

```

108 Haar_approximation_function=function(Z){
109   f=rep(0,length(Z))
110
111
112   for (i in 1:((log2(length(Z))/log2(2))-1)){
113     if (i==1){a=""}
114     else{a="+"}
115     plot(Haar_approximation_father_by_level(i,Z),type='l',ylab=a,axes=F,col="green")
116
117     f=f+Haar_approximation_father_by_level(i,Z)
118   }
119
120
121   plot(Haar_approximation_mother_by_level(i,Z),type='l',col='red',ylab="+",axes=F)
122   f=f+Haar_approximation_mother_by_level(i,Z)
123   return(f)
124 }
125
126 par(mfrow=c(4,3))
127
128 plot(1:length(Z),Haar_approximation_function(Z),axes=F,type='l',ylim=c(0,max(Z)),xlim=c(0,
129   length(Z)),col='blue',xlab="",ylab="")
130
131 par(new=T)
132 plot(1:length(Z),Z,ylim=c(0,max(Z)),axes=F,xlim=c(0,length(Z)),col='red',xlab="",
133   ylab="||")

```

Annexe D

Inversion des formes d'onde LiDAR à l'aide de l'Analyse en Composantes Principales

```
1 #####LIBRAIRIES#####
2
3 library (FactoMineR)
4 library (tree)
5 library (randomForest)
6
7 #####IMPORTATION DE LA BASE DE DONNEES#####
8
9 rm (list=ls ())
10 setwd ("/homeL/mezian/Dropbox/Morbihan-captreur1 (toutGAPD) ")
11 donnees_morbihan1<-read.csv (file="toutGAPDchampionne.csv", sep=" ", dec=".", header=F)
12
13 #####PREPARATION DONNEES#####
14 Signal_begin=rep (NA, nrow (donnees_morbihan1))
15 Signal_end=rep (NA, nrow (donnees_morbihan1))
16
17 for (i in 1:189){
18   p=donnees_morbihan1 [i, 2:201]
19   line=(5*apply (donnees_morbihan1 [i, 90:200], 1, sd))+apply (donnees_morbihan1 [i, 90:200], 1, mean)
20   b=p [p>line]
21   Signal_begin [i]=min (which (p==b [1]))
22   Signal_end [i]=max (which (p==b [length (b)]))
23 }
24
25 wext=max (Signal_end)-min (Signal_begin)-4
26
27 tableau1=donnees_morbihan1 [1, Signal_begin [1]:(Signal_begin [1]+wext)]
28
29 tableau2=donnees_morbihan1 [2, Signal_begin [2]:(Signal_begin [2]+wext)]
30
31 names (tableau2)=names (tableau1)
32
33 tableau=rbind (tableau1, tableau2)
34
35 names (tableau)=names (tableau1)
36
37 for (i in 3:189){
38   tableaux=donnees_morbihan1 [i, Signal_begin [i]:(Signal_begin [i]+wext)]
39   names (tableaux)=names (tableau1)
```

ANNEXE D. INVERSION DES FORMES D'ONDE LIDAR À L'AIDE DE L'ANALYSE EN
COMPOSANTES PRINCIPALES

```

40  tableau=rbind(tableau , tableaux)
41
42 }
43
44 names(tableau)= paste("P(t" ,1:64 ,")" ,sep="")
45 hauteur=donnees_morbihan1[,1]
46 tableau=cbind(hauteur , tableau)
47 couleur<-cut(hauteur , breaks=seq(0 ,max(hauteur)+1, by=1.5))
48
49 acpx=PCA(tableau[,c(2:65) ] , scale.unit=FALSE, ncp=length(2:65) ,graph=T)
50 tableaux=cbind(hauteur , acpx$ind$coord[,1:30])
51 tableaux=as.data.frame(tableaux)
52
53 #####PARTIE UTILE #####
54 wext=Signal_end-Signal_begin
55
56
57 #####VALIDATION CROISEE #####
58 k=5
59 folds=cvsegments(nrow(tableau) ,k)
60 result=rep(NA,nrow(tableau))
61 for (fold in 1 : k){
62   currentFold= folds[[fold]]
63
64 #####ACP #####
65   acp<-PCA(tableau[,c(2:65) ] , scale.unit=FALSE, ncp=length(2:65) ,graph=F, ind.sup=
        currentFold)
66
67
68   h=hauteur[-currentFold]
69
70   datan1=cbind(h , acp$ind$coord[,1:30])
71
72   w=wext[-currentFold]
73
74   datan1=cbind(w , datan1)
75
76   datan1 = as.data.frame.matrix(datan1)
77 #####CONSTRUCTION MODELE STATISTIQUE RLM #####
78   fit=lm(h ~Dim.1 + Dim.2 + Dim.3 ,data=datan1);
79
80
81   h=hauteur[currentFold]
82
83   datan2=cbind(h , acp$ind.sup$coord[,1:30])
84
85   w=wext[currentFold]
86
87   datan2=cbind(w , datan2)
88
89   datan2 = as.data.frame.matrix(datan2)
90 #####VALIDATION MODELE PREDICTIF #####
91   result[currentFold]= predict(fit , datan2)
92
93 }
94
95 plot(hauteur , result ,ylab="Bathymétrie estimée (m)" ,xlab="Bathymétrie mesurée (m)" ,main=titre
        ,pch=20)
96 abline(0,1 ,col='red')
97 print(cor(hauteur , result)^2)
98 mtext(paste("ACP" , "RF Step+Wext "))

```

Annexe E

Inversion des formes d'onde LiDAR à l'aide de l'analyse en ondelettes

```

1 #####LIBRAIRIES#####
2
3 library (FactoMineR)
4 library (tree)
5 library (randomForest)
6
7 #####IMPORTATION DE LA BASE DE DONNEES#####
8 rm (list=ls ())
9 setwd ("/homeL/mezian/Dropbox/Morbihan-capt eur1 (toutGAPD) ")
10 donnees_morbihan1<-read.csv (file="toutGAPDechantillonne.csv", sep=" ", dec=".", header=F)
11
12 Signal_begin=rep (NA, nrow (donnees_morbihan1))
13 Signal_end=rep (NA, nrow (donnees_morbihan1))
14 #####PREPARATION DONNEES#####
15 for (i in 1:189){
16
17     p=donnees_morbihan1 [i, 2:201]
18
19     line=(5*apply (donnees_morbihan1 [i, 90:200], 1, sd))+apply (donnees_morbihan1 [i, 90:200], 1, mean)
20
21     b=p [p>line]
22
23     Signal_begin [i]=min (which (p==b [1]))
24
25     Signal_end [i]=max (which (p==b [length (b)]))
26
27 }
28
29 wext=max (Signal_end)-min (Signal_begin)-4
30
31 tableau1=donnees_morbihan1 [1, Signal_begin [1]:(Signal_begin [1]+wext)]
32
33 tableau2=donnees_morbihan1 [2, Signal_begin [2]:(Signal_begin [2]+wext)]
34
35 names (tableau2)=names (tableau1)
36
37 tableau=rbind (tableau1, tableau2)
38
39 names (tableau)=names (tableau1)
40
41
42 for (i in 3:189){
43

```

ANNEXE E. INVERSION DES FORMES D'ONDE LIDAR À L'AIDE DE L'ANALYSE EN
ONDELETTES

```

44  tableaux=donnees_morbihan1[i,Signal_begin[i):(Signal_begin[i]+wext)]
45
46  names(tableaux)=names(tableau1)
47
48  tableau=rbind(tableau,tableaux)
49
50 }
51
52 names(tableau)=paste("P(t",1:64,")",sep="")
53
54 hauteur=donnees_morbihan1[,1]
55
56 tableau=cbind(hauteur,tableau)
57
58 #####COEFFICIENT ONDELETTES#####
59 a1=dwt(tableau[1,2:65], 'haar', (log2(length(tableau[1,2:65]))/log2(2))-1)
60
61 a=dwt(tableau[2,2:65], 'haar', (log2(length(tableau[2,2:65]))/log2(2))-1)
62
63 c=rbind(c(a1[[5]], a1[[6]]), c(a[[5]], a[[6]]) ) #2 NIVEAUX
64
65 colnames(c)=paste("c",1:4,sep="")
66
67 for (i in 3:189){
68
69   a=dwt(tableau[i,2:65], 'haar', (log2(length(tableau[i,2:65]))/log2(2))-1)
70
71   c=rbind(c,c(a[[5]], a[[6]]) )
72
73 }
74
75 k=5
76 wext=Signal_end-Signal_begin
77 datan2=cbind(hauteur, wext, c)
78 wext=Signal_end-Signal_begin
79 datan2 = as.data.frame.matrix(datan2)
80 folds=cvsegments(nrow(datan2),k)
81 result=rep(NA,nrow(datan2))
82
83 ##### VALIDATION CROISEE#####
84 for (fold in 1:k){
85   currentFold= folds[[fold]]
86   fit=randomForest(hauteur ~ c1 + c2 + c3+wext, data=datan2[-currentFold,])
87   result [currentFold]= predict(fit, datan2[currentFold,])
88 }
89
90 a=round(100* (cor(hauteur, result)^2),2 )
91 b=round(rmse(hauteur, result),2 )
92
93 plot(hauteur, result, ylab="Bathymétrie estimée (m)", pch=20, xlab="Bathymétrie mesurée (m)")
94 abline(0,1, col='red')
95
96
97 mtext(paste("Ondelette", "RF Step (4 coefficients) Wext"))
98 print(a); print(b)

```

Annexe F

Inversion des formes d'onde LiDAR à l'aide des splines cubiques

```

1
2 #####IMPORTATION DONNEES#####
3 rm(list=ls())
4 setwd("/homeL/mezian/R/WaLiD_R_122013_JeanStephane_underconstruction/lidar_foret_bresil/
5 formes_ondes_brezil/raw_waveforme/")
6 donnees_morbihan1<-read.csv(file="toutGAPDechantillonne.csv",sep=" ",dec=".",header=F)
7
8 Signal_begin=rep(NA,nrow(donnees_morbihan1))
9 Signal_end=rep(NA,nrow(donnees_morbihan1))
10 #####PREPARATION DONNEES#####
11 for(i in 1:189){
12   p=donnees_morbihan1[i,2:201]
13   line=(5*apply(donnees_morbihan1[i,90:200],1,sd))+apply(donnees_morbihan1[i,90:200],1,mean)
14   b=p[p>line]
15   Signal_begin[i]=min(which(p==b[1]))
16   Signal_end[i]=max(which(p==b[length(b)]))
17 }
18 wext=max(Signal_end)-min(Signal_begin)-4
19 tableau=c()
20 for(i in 1:189){
21   tableaux=donnees_morbihan1[i,Signal_begin[i):(Signal_begin[i]+wext)]
22   names(tableaux)=paste("P(i",1:64,")",sep="")
23   tableau=rbind(tableau,tableaux)
24 }
25
26 hauteur=donnees_morbihan1[,1]
27
28 tableau=cbind(hauteur,tableau)
29
30 #####COEFFICIENTS SPLINES CUBIQUES#####
31 U=c()
32 for(j in 1:189){
33   n=31
34   x=seq(2,65,length=n)
35   y=tableau[j,x]
36
37   ispl <- interpSpline(x, y)
38   c=c()
39   for(i in 1:(n-1)){
40
41     c=c(c, ispl$coefficients[i,c(1,2,3,4)])
42   }

```

ANNEXE F. INVERSION DES FORMES D'ONDE LIDAR À L'AIDE DES SPLINES CUBIQUES

```

43  U=rbind(U,c)
44
45  }
46
47  hauteur=donnees_morbihan1[,1]
48  u=cbind(hauteur,U)
49  w=Signal_end-Signal_begin
50  u=cbind(w,u)
51  datan1 = as.data.frame.matrix(u)
52
53  #####VALIDATION CROISEE#####
54  library(tree)
55  library(randomForest)
56  k=5
57  folds=cvsegments(nrow(datan1),k)
58  result=rep(NA,nrow(datan1))
59
60  for (fold in 1 : k){
61    currentFold= folds[[fold]]
62    fit=randomForest(hauteur~V4 + V6 + V10 + V14 + V18 + V22 + V26 +
63                    V30 + V34 + V38 + V42 + V46 + V50 + V54 + V58 + V62 + V74+w,data=datan1[-
64                    currentFold,])
65    result [currentFold]= predict(fit , datan1[currentFold,])
66  }
67
68
69
70  residu=result-hauteur
71  R2=1-(var(residu)/var(hauteur))
72  a=round(R2*100,2 )
73  b=round(rmse(hauteur , result ) ,2 )
74
75  plot (hauteur , result , xlab="Bathymétrie mesurée",ylab="Bathymétrie estimée" ,ylim=c(0,3) ,pch
76        =20 ,main="Spline cubique")
77  abline(0,1 ,col='red')

```

Annexe G

Inversion des formes d'onde LiDAR à l'aide des B-Splines cubiques

```

1 #####IMPORTATION DONNEES#####
2 rm(list=ls())
3 setwd("/homeL/mezian/Dropbox/Morbihan-captteur1(toutGAPD)")
4 donnees_morbihan1<-read.csv(file="toutGAPDechantillonne.csv",sep=" ",dec=".",header=F)
5
6
7 Signal_begin=rep(NA,nrow(donnees_morbihan1))
8 Signal_end=rep(NA,nrow(donnees_morbihan1))
9 #####PREPARATION DONNEES#####
10 for(i in 1:189){
11
12     p=donnees_morbihan1[i,2:201]
13
14     line=(5*apply(donnees_morbihan1[i,90:200],1,sd))+apply(donnees_morbihan1[i,90:200],1,mean)
15
16     b=p[p>line]
17
18     Signal_begin[i]=min(which(p==b[1]))
19
20     Signal_end[i]=max(which(p==b[length(b)]))
21
22 }
23
24 wext=max(Signal_end)-min(Signal_begin)-4
25
26
27 tableau=c()
28
29 for(i in 1:189){
30
31     tableaux=donnees_morbihan1[i,Signal_begin[i):(Signal_begin[i]+wext)]
32
33     names(tableaux)=paste("P(t",1:64,")",sep=" ")
34
35     tableau=rbind(tableau,tableaux)
36
37 }
38
39 hauteur=donnees_morbihan1[,1]
40
41 tableau=cbind(hauteur,tableau)
42
43

```

ANNEXE G. INVERSION DES FORMES D'ONDE LIDAR À L'AIDE DES B-SPLINES CUBIQUES

```

44 n=30
45 t=matrix(0,nrow(donnees_morbihan1),n)
46
47 #####COEFFICIENTS B-SPLINES #####
48
49 for(i in 1:nrow(donnees_morbihan1)){
50   i=10
51   y=t(tableau[i,2:65])
52
53   basis = create.bspline.basis(c(2,65), nbasis=n, norder=4)
54
55   fdParobj = fdPar(fdobj=basis, Lfdobj=2, lambda=0)
56   plfd = smooth.basis(2:65,y,fdParobj=fdParobj)$fd
57
58   t[i,]=t(plfd$coefs)
59
60 }
61
62 hauteur=donnees_morbihan1[,1]
63 u=cbind(hauteur,t)
64 wext=Signal_end-Signal_begin
65 u=cbind(u,wext)
66 datan1 = as.data.frame.matrix(u)
67
68 library(tree)
69 library(randomForest)
70 k=5
71 folds=cvsegments(nrow(datan1),k)
72 result=rep(NA,nrow(datan1))
73
74 #####VALIDATION CROISEE#####
75
76 for(fold in 1:k){
77   currentFold= folds[[fold]]
78   fit=randomForest(hauteur ~V3 + V4 + V5+wext, data=datan1[-currentFold,])
79   result [currentFold]= predict(fit, datan1[currentFold,])
80 }
81
82
83 plot(hauteur,result,xlab="Bathymétrie mesurée (m)",ylab="Bathymétrie estimée (m)",main="B-
      Spline",pch=20)
84 abline(0,1,col='red')
85
86 residu=result-hauteur
87 R2=1-(var(residu)/var(hauteur))
88 a=round(100*cor(hauteur,result)^2,2)
89 b=round(rmse(hauteur,result),2)

```

Annexe H

Validation croisée 5-folds

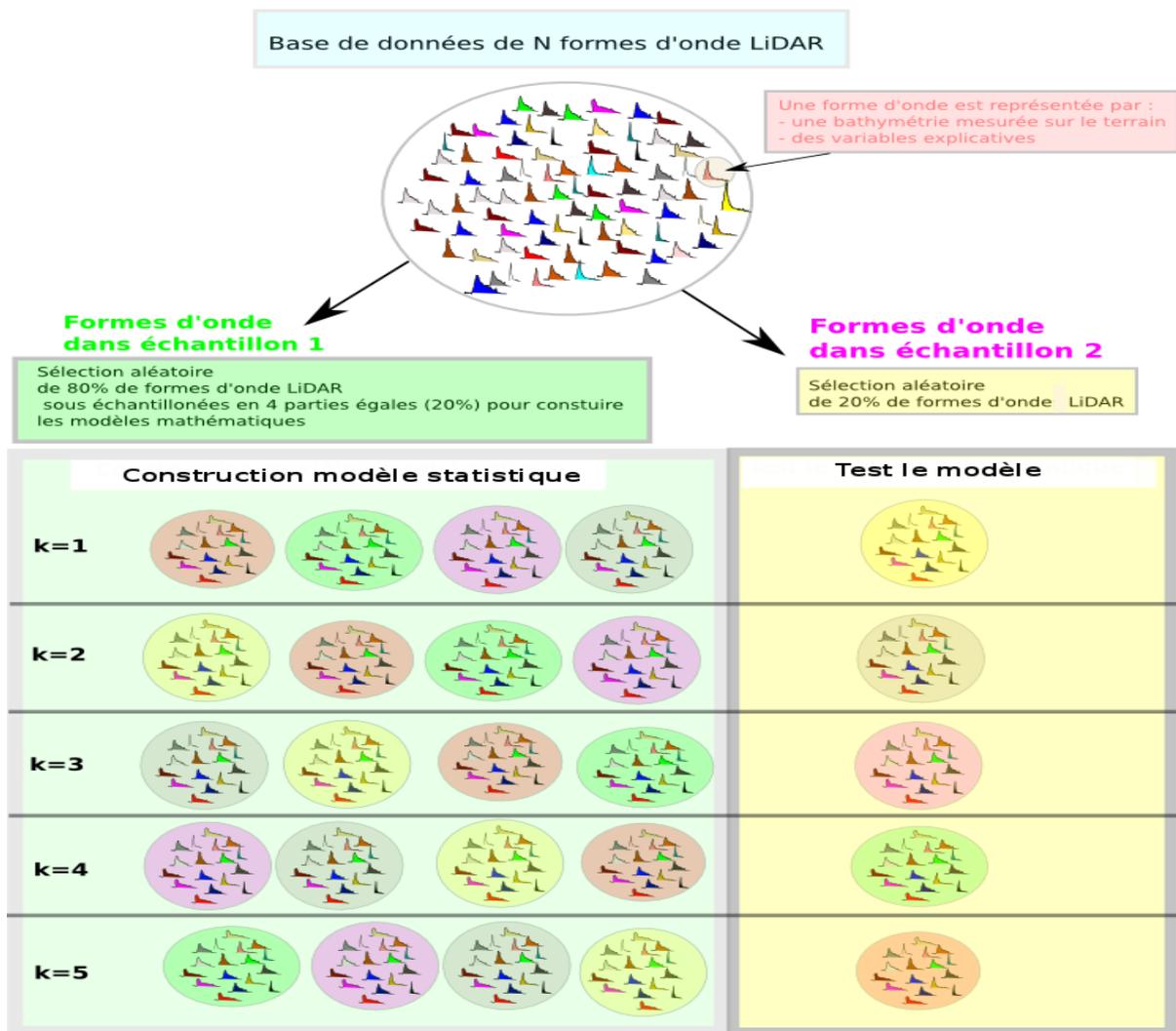


FIGURE H.1 – Schéma de la validation croisée 5-folds

Annexe I

Stepwise : Méthode de sélection automatique de variables explicatives

L'objectif de tout statisticien utilisant une régression linéaire multiple est d'aboutir, à partir d'un **ensemble de variables explicatives**, à un modèle final qui retiendrait le plus grand nombre de variables explicatives qui s'avèrent significatives dans l'explication de la variation de la variable d'intérêt Y .

La méthode "**Stepwise**" [65] est une **méthode de régression dite "pas à pas"**. A chaque étape de la procédure, on examine à la fois si une nouvelle variable explicative doit être ajoutée selon un seuil d'entrée fixé, et si une des variables déjà incluses doit être éliminée selon un seuil de sortie fixé. Cette méthode permet de retirer du modèle d'éventuelles variables qui seraient devenues moins indispensables du fait de la présence de celles nouvellement introduites. La procédure s'arrête lorsque aucune variable ne peut être rajoutée ou retirée du modèle selon les critères choisis. Les variables sont retenues ou éliminées du modèle sur la base de critères uniquement statistiques.

| Variable d'intérêt | Variables explicatives | | | |
|--------------------|------------------------|-------|-----|-------|
| Y | X_1 | X_2 | ... | X_n |

TABLE I.1 – Tableau que l'on nomme "Table"

Nous allons expliquer à présent comment faire un **stepwise** avec **R**.

```

1 reg=lm( Y ~ . , data="Table")#Regression lineaire multiple sur toutes les variables
                                explicatives du tableau "Table"
2
3 reg=step( reg)                 # Fonction Stepwise : Selection automatique des variables
                                explicatives
4
5 summary( reg)                 # Indique les variables explicatives retenues

```

Annexe J

CART

Cette méthode consiste à utiliser des variables explicatives pour diviser les individus en groupes homogènes. Les résultats se présentent sous la forme de règles de décisions qui permettent aisément de classer les individus pour la prédiction. La forme graphique des résultats est un arbre hiérarchisé.

Le principe de fonctionnement est le suivant. L'arbre commence par choisir la variable qui par ses modalités, sépare le mieux les individus de chaque classe, appelé noeud, de façon à avoir des sous populations contenant le plus possible d'individus d'une seule classe. On réitère cette opération sur chaque nouveau noeud jusqu'à ce que la séparation des individus ne soit plus possible, dans le cas où on a peu d'individus par classes ou bien un nombre d'individu en dessous d'un seuil fixé.

Annexe K

Root Mean Square Error (RMSE)

L'erreur quadratique moyenne (RMSE) s'écrit :

$$RMSE = \sqrt{\frac{\sum_{j=1}^n (y_j - \hat{y}_j)^2}{n}} \quad (\text{K.1})$$

- y_j est la valeur mesurée
- \hat{y}_j est la valeur estimée
- n est le nombre d'observations

Annexe L

Les composantes principales

Les nouveaux axes, appelés composantes principales, sont construits de la façon suivante :

- la **première composante principale** est la direction de "**plus grand allongement**" du nuage (Figure L.3). Lorsque on projette les points du nuage sur cet axe, leurs projections sont plus dispersées qu'elles ne le seraient sur n'importe quel autre axe. La première composante principale est donc l'axe selon lequel est préservé, par projection, le maximum de la dispersion initiale des points du nuage. La nouvelle variable PC_1 (ou la première composante principale) est le caractère selon lequel les individus se différencient le plus.

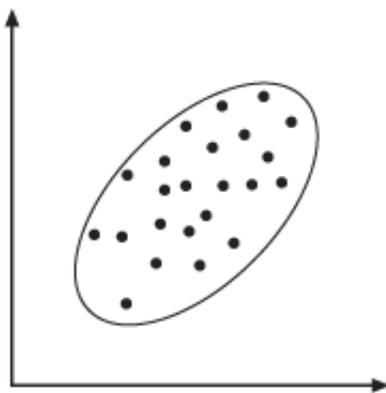


FIGURE L.1 – Nuage de points. On cherche la direction de "plus grand allongement" du nuage
- Référence : [80]

allongement
maximal

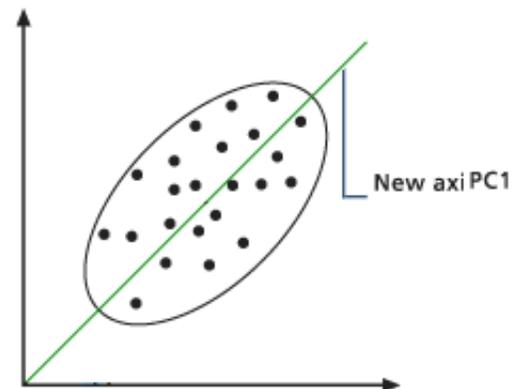


FIGURE L.2 – La première composante principale (PC : " Principal component ") - Référence : [80]

- la **deuxième composante principale** est la **deuxième direction d'allongement du nuage**, c'est-à-dire celle qui explique, après le première axe, le maximum de l'inertie résiduelle. De plus, le deuxième axe est choisi orthogonal au premier, ce qui traduit le fait que la deuxième composante principale est **non corrélée à la première**.

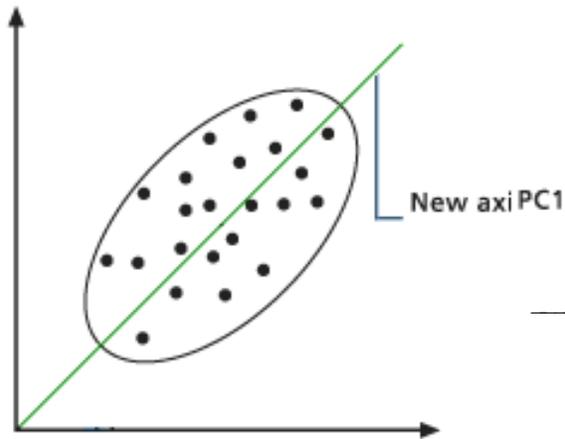


FIGURE L.3 – On cherche la direction avec le maximum de l'inertie résiduelle de sorte que PC_2 soit perpendiculaire (non corélées) à PC_1 - Référence : [80]

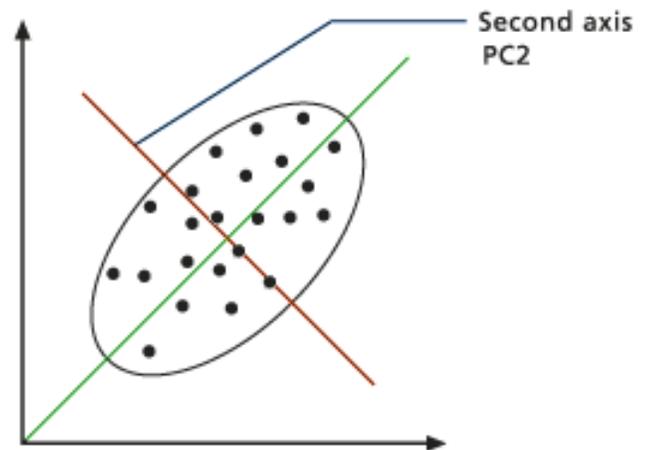


FIGURE L.4 – La deuxième composante principale (PC : " Principal component ") - - Référence : [80]

- et nous continuons, jusqu'à avoir remplacé les p anciens axes par p nouveaux axes (les composantes principales), portant des parts décroissantes de la dispersion initiale et dont les premières nouvelles variables suffisent souvent à décrire une grande partie de l'information du tableau initial (Figure 8.1).

Le pourcentage d'inertie¹ expliquée par une composante principale permet d'évaluer en quelque sorte la quantité d'information recueillie par cet axe.

1. ou variance

Annexe M

Fonction d'échelle père ϕ et l'ondelette mère ψ

La **fonction d'échelle père** ϕ et **l'ondelette mère** ψ sont définis de la manière suivante :

Soit $\phi = 1$ sur $[0, 1]$, et $\psi(t) = \begin{cases} -1 & \text{si } t \in [0; \frac{1}{2}[\\ 1 & \text{si } t \in [\frac{1}{2}; 1[\end{cases}$

Et **l'ondelette mère** $\psi_{k,w}$ pour $k \geq 0$ et $0 \leq w \leq 2^k - 1$, on pose $\psi_{k,w}(t) = 2^{k/2}\psi(2^k t - w)$:

$$\psi_{k,w}(t) = \begin{cases} -1 & \text{si } t \in [w 2^{-k}, (w + \frac{1}{2}) 2^{-k}[\\ 1 & \text{si } t \in [(w + \frac{1}{2}) 2^{-k}, (w + 1) 2^{-k}[\end{cases} \quad (\text{M.1})$$

avec k le niveau de résolution, w le facteur de translation et $N = 2^k$ le nombre de points du signal.

Annexe N

Les coefficients des splines cubiques

Pour plus d'informations sur les splines cubiques voir [27].

Soit S de classe C^2 sur I , cubique sur chaque intervalle I_i et coïncide avec f aux points t_i et t_{i+1} . Soit p_i la restriction de S à l'intervalle I_i . On peut chercher p_i sous la forme :

$$p_i(t) = a_i(t - t_i)^3 + b_i(t - t_i)^2 + c_i(t - t_i) + d_i \quad i = 0, \dots, n - 1 \quad (\text{N.1})$$

Cherchons les coefficients a_i, b_i, c_i et d_i

Les fonctions f et p_i coïncide aux points t_i et t_{i+1} , on a :

$$p_i(t_i) = f(t_i) \quad \text{et} \quad p_i(t_{i+1}) = f(t_{i+1}) \quad (\text{N.2})$$

$$\boxed{d_i = f(t_i)} \quad (\text{N.3})$$

$$a_i h^3 + b_i h^2 + c_i h + d_i = f(t_{i+1}) \quad (\text{N.4})$$

avec $h = t_{i+1} - t_i$.

On peut exprimer les inconnues a_i, b_i et c_i (d_i étant connus) en fonction des dérivées de la fonction. On note

$\alpha_i = p''(t_i)$. En écrivant la continuité de la dérivée seconde aux points t_{i+1} , on a :

$$p_i''(t_{i+1}) = p_{i+1}''(t_{i+1}) \quad (\text{N.5})$$

avec

$$p_i'(t_{i+1}) = 3a_i h^2 + 2b_i h + c_i \quad (\text{N.6})$$

$$p_i''(t_{i+1}) = 6a_i h + 2b_i \quad (\text{N.7})$$

$$\alpha_i = p_i''(t_i) = 6a_i(t_i - t_i) + 2b_i = 2b_i \quad (\text{N.8})$$

$$\boxed{b_i = \frac{\alpha_i}{2}} \quad (\text{N.9})$$

$$p_i''(t_{i+1}) = 6a_i h + 2b_i = \alpha_{i+1} \quad (\text{N.10})$$

$$6a_i h + \alpha_i = \alpha_{i+1} \quad (\text{N.11})$$

$$\boxed{a_i = \frac{\alpha_{i+1} - \alpha_i}{6h}} \quad (\text{N.12})$$

Substituant les relations (N.9) et (N.12) dans (N.4), on obtient :

$$f_{i+1} = \frac{(\alpha_{i+1} - \alpha_i)h^2}{6} + \frac{\alpha_i h^2}{2} + c_i h + f_i \quad (\text{N.13})$$

d'où l'expression de c_i

$$\boxed{c_i = \frac{f_{i+1} - f_i}{h} - \frac{(2\alpha_i + \alpha_{i+1})h}{6}} \quad (\text{N.14})$$

En écrivant maintenant la continuité de la dérivée première au point t_i , on obtient

$$3a_{i-1}h^2 + 2b_{i-1}h + c_{i-1} = c_i \quad (\text{N.15})$$

En réécrivant les a_i, b_i, c_i figurant dans (N.15) par leurs expressions données par (N.9), (N.12) et (N.14) on obtient n-1 équations qui expriment une relation de récurrence entre α_{i-1}, α_i et α_{i+1} :

$$h(\alpha_{i-1} + 4\alpha_i + \alpha_{i+1}) = \frac{6}{h}(f_{i-1} - 2f_i + f_{i+1}) \quad i = 1, \dots, n-1 \quad (\text{N.16})$$

A ces n-1 équations, il faut ajouter deux autres conditions : On peut par exemple poser $\alpha_0 = 0$ et $\alpha_n = 0$, on parle alors de spline naturelle. Le vecteur $\alpha = (\alpha_1, \dots, \alpha_{n-1})^T$ est solution du système tridiagonal

$$A\alpha = b \quad (\text{N.17})$$

avec :

$$A = h \begin{pmatrix} 4 & 1 & 0 & \dots & \dots & 0 \\ 1 & 4 & 1 & \dots & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & 1 & 4 & 1 \\ 0 & 0 & \dots & 1 & 4 & \end{pmatrix}$$

$$b = \frac{h}{6} \begin{pmatrix} f_0 - 2f_1 + f_2 \\ \vdots \\ f_{i-1} - 2f_i + f_{i+1} \\ \vdots \\ f_{n-2} - 2f_{n-1} + f_n \end{pmatrix}$$

La matrice A est inversible, car elle est dite à diagonale strictement dominante. Connaissant les α_i , on détermine les p_i par les relations (N.12), (N.9), (N.14) et (N.3) et la spline est déterminé sur chaque intervalle.

Démarches :

- On construit la matrice A
- On construit le second membre b
- On résout le système linéaire (N.17) pour déterminer les α_i
- On calcul les coefficients a_i, b_i, d_i et c_i pour $i = 0, \dots, n-1$
- On calcul les $p_i(i)$ sur chaque interval $[t_i, t_{i+1}]$ pour $i = 0, \dots, n-1$

Voici un code qui m'as permis d'approcher une fonction quelconque par un nombre n de polynômes cubiques par morceaux.

```

1 library (SoDA)
2
3 # fonction a interpoler
4 f=function (x){
5   f=sin(4*pi*x)
6   return (f)
7 }
8
9 n=10;
10
11 x=(0:n)/n
12 h=1/n
13 fx=f(x);
14
15 b=-2*fx [2:n]+fx [1:(n-1)]+fx [3:(n+1)];
16 b=6*b/h;
17
18 A=h*tridiag (rep (4,n-1),1,1)
19 alpha=solve(A,b);
20 alpha=c(0,alpha,0);
21
22 coefficient=rep (NA,4*n)
23 j=0
24 for (i in 1:n){
25
26   a=(alpha [i+1]-alpha [i])/(6*h);
27   b=alpha [i]/2;
28   c=((fx [i+1]-fx [i])/h) - ( (2*alpha [i]+alpha [i+1])*h ) /6;
29
30
31   coefficient [j+1]=a
32   coefficient [j+2]=b
33   coefficient [j+3]=c
34   coefficient [j+4]=fx [i]
35   j=j+4
36
37   Ii=seq(x [i],x [i+1],length=20)
38
39   fi=f(Ii);
40   u=(Ii-x [i])
41   Si=a*u^3+b*u^2+c*u+fx [i];
42   plot (Ii,fi,xlim=c(0,1),ylim=c(-1,1),col='gray',pch=20,ylab="f(x)",xlab="x");
43   par(new=T)
44   plot (Ii,Si,xlim=c(0,1),ylim=c(-1,1),type='l',col=i,main="Spline cubique",xlab="",ylab="")
45   ;
46   par(new=T)
47 }
48 abline (v=x,pch=2)

```

Bibliographie

- [1] A. Boudahoui Effet d'une géométrie de fond irrégulière sur l'estimation bathymétrique à partir de formes d'onde LiDAR, 2013
- [2] Alain Batailly, Note technique Opérations élémentaires sur courbes splines en 2D, 2012.
- [3] Guenther G.C., Airborne Laser Hydrography, system design and performance factors, NOAA Professional Paper Ser. NOS1 (National Oceanic and Atmospheric Administration, Rockville, Md., 1985
- [4] Tulldahl H.M., Steinvall K.O., Simulation of sea surface wave influence on small target detection with airborne laser depth sounding, *Appl. Opt* 42, 2462-2483, 2004
- [5] Mobley C.D., *Light and Water : Radiative Transfer in Natural Waters*, Academic, San Diego, Calif., 1994
- [6] Austin R.W. and T.J. Petzold, 1986. Spectral dependence of the diffuse attenuation coefficient of light. *Ocean water Opt. Eng.*, vol. 25, pp. 471-479.
- [7] Beckmann P. and Spizzochino A. 1963. The scattering of electromagnetic waves from rough Surfaces. *IEEE transactions on antennas and propagation*, vol. 56, pp. 2043-2050.
- [8] Blinn J.F. 1977. Model of Light Reflection for Computer Synthesized Pictures. In *Proceedings SIGGRAPH'77*, vol. 11, pp. 192-198.
- [9] Boris J., Stilla U., 2006. Range determination with waveform recording laser systems using a Wiener Filter. *ISPRS Journal of Photogrammetry and Remote Sensing* vol. 61, pp. 95-107
- [10] Brenner, A.C., Zwally, H.J., Bentley, C.R., Csatho, B.M., Harding, D.J., Hofton, M.A., Minister, J.B., Roberts, L.A., Saba, J.L., Thomas, R.H., Yi, Y., 2003. Geoscience Laser Altimeter System (GLAS) — derivation of range and range distributions from laser pulse waveform analysis for surface elevations, roughness, slope, and vegetation heights.
- [11] Bricaud, A., Morel, A. and Prieur, L. 1981. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains *Limnology and Oceanography*, vol. 26, p. 43-53.
- [12] Bricaud, A. and Stramski, D. 1990. Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter : a comparison between the Peru upwelling area and the Sargasso Sea. *Limnology and Oceanography*, vol. 35, pp. 563-582.
- [13] Bukata, R., Jerome, J., Kondratyev, K., et Pozdnyakov, D. 1995. *Optical properties and remote sensing of inland and coastal waters*, CRC press, Florida.
- [14] Cipollini, P. and Corsini, G. 1994. The effect of yellow substance on pigment concentration retrieval using 'blue to green' ratio. *Proceeding of IEEE International Geoscience and Remote Sensing Symposium, IGARSS-94*, pp. 772-777.
- [15] Cook R.L. and Torrance K.E. 1982. A Reflectance Model for Computer Graphics. *ACM Transactions on Graphics*, vol. 1, pp. 7 - 24.
- [16] Cox, G. and Munk, W. 1954. Measurement of the roughness of the sea surface from photographs of the sun's glitter, vol 11, pp. 838 - 850.

- [17] Chisheng Wang and al. A Comparison of Waveform Processing Algorithms on Single-Wavelength LiDAR Bathymetry 2014
- [18] Eisma D., 1993. *Suspended Matter in the Aquatic Environment*, Springer-Verlag, Berlin.
- [19] Feigels J. 1992. Lidars for oceanological research : criteria for comparison, main limitations, perspectives. *Ocean Optics XI, SPIE 1750*, pp. 473–484.
- [20] Gaston G., 2003. Etude comparative de fonctions de réflectance bidirectionnelle calculées par méthode de Monte-Carlo et de leurs modèles théoriques. Master's thesis, ENSMA.
- [21] Gilvear D., Hunter P. and Higgins T. 2007. An experimental approach to the measurement of the effects of water depth and substrate on optical and near infra-red reflectance : A field-based assessment of the feasibility of mapping submerged instream habitat. *International Journal of Remote Sensing*, vol. 28, pp. 2241–2256.
- [22] Guenther G. C. and Thomas R. W. L., 1984. "Prediction and correction of propagation induced depth measurement biases plus signal attenuation and beam spreading for airborne laser hydrography," NOAA Tech. Rep. NOS 106, Washington 106 pp.
- [23] Guenther G.C., 1985. Airborne Laser Hydrography, system design and performance factors, NOAA Professional Paper Ser. NOS1 (National Oceanic and Atmospheric Administration), 395 pp.
- [24] Guenther G.C., LaRocque P. and Lillycrop W. 1994. Multiple surface channels in SHOALS airborne lidar. *SPIE : Ocean Optics XII 2258*, pp. 422-430.
- [25] Guenther G.C., Cunningham A.G., LaRocque P.E., and Reid D.J. 2000. Meeting the accuracy challenge in airborne lidar bathymetry, *Proceedings of EARSeL-SIG-Workshop LiDAR, Dresden/FRG*, vol. 1, pp. 6 – 14.
- [26] Hale M. and Query R., 1973. Optical Constants of Water in the 200-nm to 200-Mm Wavelength Region. *Applied optics*, vol. 12.
- [27] Ionut Danaila, Pascal Joly, Sidi Mahmoud Kaber, Marie Postel, *Introduction au calcul scientifique par la pratique*, 2005
- [28] Hochberg E.J., Atkinson M.J. and Andrefouet S. 2003. Spectral reflectance of coral reef bottom-types worldwide and implications for coral reef remote sensing. *Remote Sensing of Environment*, vol. 85, pp. 159–173.
- [29] Ivanoff A. 1975. *Introduction à l'océanographie. Propriétés physiques et chimiques des eaux de mer*. Librairie Vuibert, Paris, Tome II, 338 p.
- [30] Jerlov N.G., 1976. *Marine Optics*, Elsevier Scientific Publishing Co., Amsterdam, vol 14 pp 231.
- [31] Kirk J.T.O. 1983. *Light and Photosynthesis in Aquatic Ecosystems*, Cambridge Univ. Press, Cambridge, 401 pp.
- [32] Legleiter J., Roberts A. and Rick L. 2009. Spectrally based remote sensing of river bathymetry. *EARTH SURFACE PROCESSES AND LANDFORMS*, vol. 34, pp. 1039–1059.
- [33] Lesaignoux A. 2006. Modélisation et simulations de trains d'ondes LiDAR "vert" : application à la détection de faibles lames d'eau en rivière, Master's Thesis, UMIL.
- [34] Lyzenga D.R. 1978. Passive remote-sensing techniques for mapping water depth and bottom features. *Applied Optics*, vol. 17, pp. 379–383.
- [35] Mallet C., Bretar F. 2009. Full-waveform topographic lidar : State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, pp. 1-16.
- [36] McLean J. W. 1990. "Modeling of ocean wave effects for lidar remote sensing," in *Ocean Optics X*, pp. 480–491.
- [37] Michael Tonon - IGN France International - *Introduction à la télédétection*
- [38] Mobley C.D. 1994. *Light and Water : Radiative Transfer in Natural Waters*, Academic, San Diego, Calif.

- [39] Morel A. 1974 Optical properties of pure water and pure sea water. In Jerlov, N.G. and Steemann Nielsen, E. (réd.) *Optical aspects of oceanography*. Academic Press, London, p. 1-24.
- [40] Morel A. and Bricaud A., 1981. Some theoretical results concerning optics of phytoplankton with special reference to remote sensing," in *Oceanography from space*, pp.313-327
- [41] Maritorena S., Morel A. and Gentili B. 1994. Diffuse-reflectance of oceanic shallow waters Influence of water depth and bottom albedo. *Limnology and Oceanography*, vol. 39, pp. 1689–1703.
- [42] Paul et al. 1982. *Dictionnaire de télédétection aérospatiale*. Paris, Masson, 236 pages.
- [43] Petri K.J. 1977. Laser Radar Reflectance of Chesapeake Bay Waters as a function of wind speed. *IEEE transaction in geoscience electronic*, vol. 15.
- [44] Petzold T.J. 1972. Volume scattering functions for selected ocean waters, Scripps Institution of oceanography, Visibility laboratory, San Diego, Calif., 79 pp.
- [45] Phong B. 1975. Illumination for Computer Generated Pictures. *Communication of the ACM*, 311–317.
- [46] Pope, R.M. and Fry, E.S. 1997. Absorption spectrum of pure water ; Integrating cavity measurements. *Applied Optics*, vol. 36, pp. 8710-8723.
- [47] Prieur L. and Sathyendranath S. 1981. An optical classification of coastal and oceanic waters based on the specific spectral absorption curves of phytoplankton pigments, dissolved organic matter, and other particulate materials, *Limnology and Oceanography*, vol. 26, pp. 671-689.
- [48] RAMSAY (J.) and al. , *Functional Data Analysis with R and MATLAB*, Springer, 2005.
- [49] Roesler S., Perry J. and Carder L. 1989. Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters, *Limnology and Oceanography* vol. 34, pp. 1510-1523.
- [50] Sathyendranath S. and Morel A. 1983. Light emerging from the sea interpretation and uses in remote sensing. *Remote Sensing Applications in Marine Science and Technology*. NATO ASI Series, Reidel Publish, Dordrecht, pp. 329-357. Sathyendranath S., Prieur L. and Morel A. 1989. A three-component model of ocean colour and its application to remote sensing of phytoplankton pigments in coastal waters," vol. 10, pp. 1373-1394.
- [51] Savitsky A. and Golay M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, vol. 36, pp. 1627– 1639.
- [52] Schmitz-Pieffer A., Viehoff T. and Grassl H. 1990. Remote sensing of coastal waters by airborne lidar and satellite radiometer ; Part 2 : Measurements. *International Journal of Remote Sensing*, vol. 11, pp. 2185-2204.
- [53] Steiner J., Termonia Y., and Deltour J. 1972. Comments on smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry*, vol. 44, pp. 1906–1909.
- [54] Steinvall K. O. 2000. Effects of target shape and reflection on laser radar cross sections, *Optical Society of America* pp. 4381 – 4391.
- [55] Steinvall K. O., Koppari K. R., and Karlsson M. 1994. Airborne laser depth sounding. System aspects and performance. In *Ocean Optics XII*, SPIE 2258, pp. 392–412.
- [56] Smith R.C. and Baker K., 1981. Optical properties of the clearest natural waters, *Applied Optics*, vol. 20, pp. 177-184.
- [57] Thomas, R.W.I. et Miller M. 1980. Personal communication.
- [58] Tulldahl H. M. and Steinvall K. O. 1999. Analytical Waveform Generation from Small Objects in Lidar Bathymetry. *Applied Optics*, vol. 38, pp. 1021-1039. Tulldahl H. M. and Steinvall K. O. 2004. Simulation of sea surface wave influence on small target detection with airborne laser depth sounding, *Applied Optics* vol. 42, pp. 2462 – 2483.

- [59] Wagner W., Ullrich A., Ducic V., Melzer T., Studnicka N. 2006. Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner. *IS - PRS Journal of Photogrammetry and Remote Sensing*, vol. 60, pp. 100–112.
- [60] Wang C. 2007. Using airborne bathymetric lidar to detect bottom type variation in shallow waters, *Remote Sensing of Environment* vol. 106, pp. 123–135.
- [61] Hani Abdallah 2012 Le développement de LiDAR satellitaire multifonctions. Analyse exploratoire du potentiel de capteurs LiDAR pour le suivi altimétrique et bathymétrique des surfaces en eau continentales et côtières
- [62] W. Marshall and B. Burk, "Received Optical Power calculations for Optical Communications Link Performance Analysis" TDA Progress Report 42-87, vol July-September 1986 , Jet Propulsion Laboratory, Pasadena, California, pp 32-40 November 15,1986.
- [63] H. Michael Tulldahl and K. Ove Steinvall Simulation of sea surface wave influence on small target detection with airborne laser depth sounding 1999.
- [64] K. Schulmeister, D. Sliney, J. Mellerio, H. Hödlmoser, G. Sonneck Human Risk Analysis Simulator for Space Lidars, Austrian Research Centers Technical Note 1, v3.1
- [65] http://campus.isped.u-bordeaux2.fr/PLEIADE/PV3/ASPX/PV2_VPS.aspx?3792
- [66] <http://www.astronautique.wikibis.com/teledetection.php>
- [67] Notions fondamentales de la télédétection : http://ocw.ncu.edu.tw/ocwdata/62004/62004-fundamentals_of_remote_sensing/tutorial/chap1/c1p1f.html
- [68] <http://eod.uclouvain.be/fr/profs/classroom-fr.pdf>
- [69] http://en.wikipedia.org/wiki/Remote_sensing
- [70] <http://fr.wikipedia.org/wiki/Bathymetrie>
- [71] <http://calamar.univ-ag.fr/uag/physique>
- [72] http://www.teledetection.fr/component/option,com_glossary/func,display
- [73] http://smc.cnes.fr/CALIPSO/GP_mission.htm
- [74] <http://www.coptercraft.com/aerial-uav-lidar-systems/>
- [75] <http://www.futura-sciences.com/magazines/matiere/infos/dossiers/d/physique-milieu-marin-proprietes-physiques-416/page/6/>
- [76] <http://hal-agroparistech.archives-ouvertes.fr/docs/00/58/12/99/PDF/MT2010-PUB00029299.pdf>
- [77] <http://theses.ulaval.ca/archimede/fichiers/25603/ch02.html>
- [78] http://www.tankonyvtar.hu/en/tartalom/tamop425/0032_recizios_mezogazdasag/ch02s04.html
- [79] http://xlim-sic.labo.univ-poitiers.fr/rView/Fichiers/documentation_techinique.pdf
- [80] <http://mathinfo.univ-reims.fr/image/siRendu/Documents/2004-Chap5-BRDF.pdf>
- [81] http://webhelp.esri.com/arcgisdesktop/9.1/printBooks_topics.cfm?pid=3042