

---

# STATISTIQUES OBTENUES A PARTIR DES DONNEES DE LIDAR SPATIAL

## 2.1 Méthode d'analyse

Dans cette première partie de l'étude, nous allons nous focaliser sur la restitution des paramètres de structure des nuages et des aérosols. Un premier algorithme a été développé pour retrouver les fonctions de densité de probabilité des sommets des structures nuageuses et des aérosols. Afin de répondre au questionnement concernant la définition des structures, deux approches ont été utilisées, qui permettent de considérer l'atmosphère soit avec une vision unidimensionnelle caractéristique des mesures instantanées, soit en terme de structure, plus proche d'une vision bi-dimensionnelle classique en imagerie spatiale.

Deux jeux de données lidar satellite sont accessibles afin de construire et tester les performances de notre algorithme :

- i) Les observations lidar effectuées à partir de la navette spatiale Discovery durant la mission LITE (Lidar In-space Technology Experiment, NASA, mission STS-64), qui s'est déroulée entre le 10 et le 20 septembre 1994).
- ii) Les données obtenues à partir de l'instrument GLAS (Geoscience Laser Satellite) sur le satellite polaire ICESat lancé en décembre 2002 également par la NASA.

Ces missions, bien que limitées dans le temps (10 jours pour LITE, un peu plus d'un mois pour GLAS si on se limite au meilleur jeu de données), nous donnent la possibilité d'estimer la distribution spatiale des nuages et des aérosols sous la trace du satellite avec une haute résolution spatiale. Les données de LITE n'ont pas fait l'objet d'une analyse en terme de produit tels que

l'altitude du sommet des nuages, comme c'est le cas pour GLAS. Nous avons donc été conduits à développer un algorithme d'analyse spécifique pour restituer les paramètres recherchés.

Dans une première partie, les méthodes existantes sont présentées. Les étapes de notre méthodologie sont ensuite expliquées. Les points critiques de l'algorithme font l'objet de plusieurs études de cas, à partir de l'utilisation des données GLAS. L'algorithme est dans un deuxième temps adapté au jeu de données LITE. Les résultats obtenus grâce à notre algorithme sur ces deux jeux de données sont ensuite présentés et comparés entre eux. Le comportement similaire des deux algorithmes étant prouvé, les climatologies nuageuses obtenues grâce aux données actives seront à terme comparées aux climatologies passives de ISCCP prises sous la trace de LITE (voir chapitre 3). Les différences sont alors discutées en terme de biais et d'erreur.

## 2.3.1 Revue de l'existant

Afin d'effectuer le traitement des données lidar, de nombreux algorithmes de seuillages ont déjà été développés par le passé.

On rappelle ici que, bien que la technologie lidar soit apparue très tôt dans le début des années 60 (plus exactement en 1962, voir l'historique sur la naissance du lidar contenue dans la partie introductive), le développement d'algorithmes permettant l'automatisation des tâches quant au traitement des données lidar n'est arrivé que bien plus tard, à la moitié des années 1980, lorsque les performances des outils informatiques le permirent (voir chapitre introductif sur l'historique des lidar, Section 1.3).

Il existe principalement deux approches basiques afin d'extraire les nuages des données lidar : la méthode par calcul de la dérivée [Pal et al., 1992], et la méthode de seuillage, telle que décrite par [Winker et Vaughan, 1994], [Clothiaux et al., 1998] et [Chazette et al, 2001].

Alors que la méthode de dérivation peut être employée avec des données lidar brutes non calibrées, et cela quelle que soit la longueur d'onde, elle ne peut souvent être utilisée que pour des systèmes lidar pointé vers le ciel. Par contre, la méthode de seuillage peut être utilisée avec les mêmes résultats aussi bien dans le cas des systèmes lidar tourné vers le ciel ou vers le sol.

Les méthodes de restitutions de structures utilisées dans le cadre du traitement des données lidar satellites des missions LITE, GLAS/ICESat et CALIOP/CALIPSO feront usage, par conséquent, de la seconde méthode.

Un algorithme de seuillage des données LITE a été développé dans le cadre de la préparation de la mission LITE, afin d'extraire le produit sommet/base de nuages, et fournir le produit aérosols. Cette méthode, appelée LITE Cloud Finding Algorithm (CFA) bien que très aboutie dans son principe n'as pas été appliquée aux données LITE. Elle a cependant servi de base de réflexion à la mise en place d'une méthode en vue du traitement des données lidar de CALIPSO (voir l'ATBD de CALIPSO, [Vaughan, 1999]).

L'instrument CALS (The Cloud and Aerosol Lidar system), volant sur l'avion de haute altitude ER-2 de la NASA, a été employé dans de multiples campagnes de mesures autour du monde. Un algorithme d'extraction a été développé afin d'extraire de ces données, la hauteur des structures nuageuses, l'épaisseur et l'émittance des nuages optiquement faible (cirrus) [Spinhirne and Hart, 1990], et de la hauteur de la couche limite atmosphérique [Palm and Spinhirne, 1987, 1998]. L'expérience qui a été retirée du traitement de ces données a permis l'élaboration d'un algorithme d'extraction des données de la mission lidar GLAS/ICESat (voir l'ATBD de GLAS, [Palm, 2001]).

La méthode de seuillage que nous avons développé est la première méthode de seuillage à avoir été testée dans le cas de données lidar spatiales. Cette méthode, après avoir été validée sur les données lidar LITE et GLAS, a pour objectif d'être dans le futur directement appliquée au cas des données lidar de la mission CALIPSO.

## 2.3.2 Présentation de la méthode utilisée

Comme nous l'avons décrit dans la section précédente, de nombreuses méthodes ont été précédemment développées afin d'effectuer le traitement des données lidar sol et avion.

Avec la migration de la technologie lidar vers les applications spatiales, les méthodes de traitement des données lidar ont été adaptées aux caractéristiques des instruments satellites.

Les premières données spatiales lidar furent constituées par celles de la mission spatiale LITE. Les données de cette dernière mission ont été fournies par la NASA à la communauté scientifique sous forme de données brutes (signal de retrodiffusion). Aucun algorithme de seuillage

n'a été développé dans le cadre de cette mission par la Science Team de LITE afin d'effectuer le traitement de ces données.

En conséquence, il a été nécessaire de développer nous mêmes un algorithme de seuillage robuste afin d'exploiter les données de cette mission pionnière.

Cet algorithme de seuillage a par conséquent été conçu avec l'expérience que nous avons au préalable du traitement des données lidar sol et avion. La science Team de GLAS a développé en parallèle de notre travail, un algorithme de seuillage de ses données lidar. A la différence de notre algorithme, l'algorithme de la Science Team de GLAS repose sur le seuillage du signal brut, alors que notre algorithme repose sur le seuillage du signal de variance.

Nous avons gardé comme idée sous jacente, lors de la construction de cet algorithme, de faire que ce dernier reste modulable. En effet, tenant compte de la croissance du nombre de missions satellites comportant un instrument Lidar (voir déjà les missions LITE, GLAS, MOLA, CALIPSO au moment de l'écriture), notre algorithme devra donc pouvoir s'adapter de manière générique aux différentes caractéristiques de ces missions.

Une première étape de l'algorithme consistera donc dans la mise en forme des données en vue de l'extraction qui sera effectuée dans la deuxième étape. Le caractère modulable de l'algorithme résidera grandement dans la manière de renseigner les paramètres fondamentaux nécessaires aux traitements des données. Cette partie comporte dans les faits peu de difficultés.

La seconde étape constituant le seuillage et l'identification des structures nuageuses et aérosols comporte en tant que tel beaucoup plus de difficultés. En effet, même si cette partie repose sur la seule opération de seuillage du signal de variance, certains points critiques de l'algorithme devront néanmoins être étudiés de façon à s'adapter à la nature des données. Les principales difficultés sont issues des différences observées quant à :

- **La résolution des données lidar utilisées** : 15 m sur la verticale pour les données de la mission LITE, pour 76.8 m en ce qui concerne les données de la mission ICESat/GLAS).
- **Le rapport signal à bruit (SNR) rencontré**: Les différences dans les caractéristiques optiques de l'instrument utilisé, l'altitude du satellite, son angle de visée expliquent les différences observées dans la moyenne de l'énergie émise et reçue, mais aussi dans la valeur de la variance du signal. Par conséquent, le rapport SNR peut varier grandement d'un type de données à l'autre.

Ces différences devront être prises en compte dans les seuils et la taille des filtres qui seront appliqués au sein de l'algorithme. Dans le cas de chaque traitement de données, une étude de sensibilité devra être au préalable effectuée.

La troisième étape de l'algorithme visera à analyser de manière statistique les sorties de l'extraction et visualiser l'ensemble des résultats. Cette dernière étape sera entièrement indépendante de la base de données utilisée.

L'étude des différentes étapes qui constitue notre méthode d'extraction est présentée dans la suite de ce chapitre.

### 2.3.3 Détection de structures

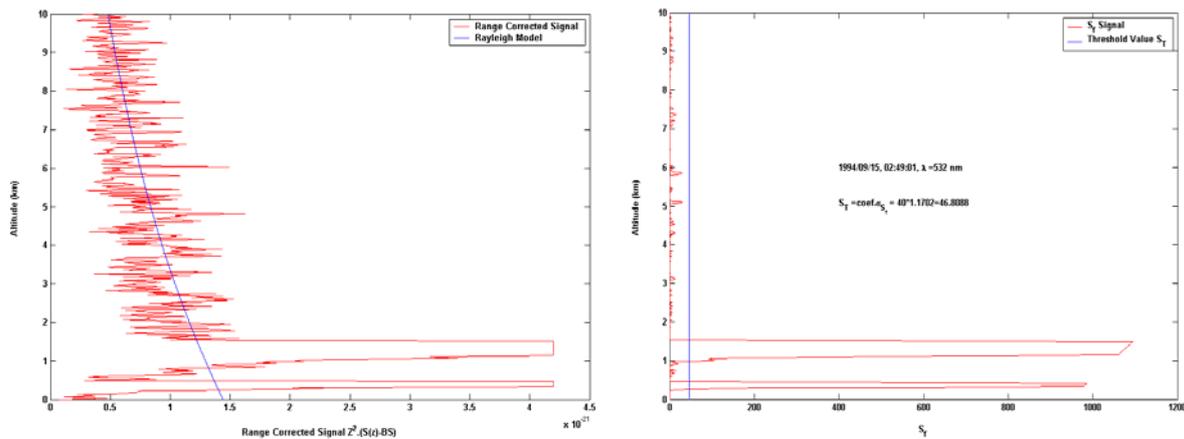
L'analyse statistique développée par Chazette et al. [2001] pour détecter des nuages à partir de signaux lidar simulés a été adaptée afin d'identifier les couches nuageuses et de déterminer leurs limites géométriques sur la verticale. La méthode est basée sur le calcul et le seuillage de la variance du signal lidar correspondant au signal brut corrigé de l'angle solide (voir la **Figure 2.1.1**).

On définit pour cela une fenêtre glissante qui est appliquée à chaque profil lidar. Ce filtrage passe-bas ne faisant appel à aucune normalisation, et pouvant être appliqué tir à tir, il permet l'identification des couches atmosphériques de manière fiable, avec une résolution sur l'horizontale identique à celle de l'échantillonnage lidar. Il peut être par principe même appliqué à toute mesure active lidar ou radar. Son avantage est de s'affranchir de la détermination du signal moléculaire (qui varie peu), ce qui permet une adaptation facile à différentes longueurs d'onde.

Cette étape nous permet d'éliminer la contribution de la diffusion moléculaire. Les profils verticaux de variance relative  $S_f$  du signal corrigé de l'angle solide  $S$ , au point d'altitude donné  $k$ , sont calculés dans un premier temps en utilisant une fenêtre de filtrage adaptée comprenant  $(2n+1)$  points:

$$S_f[k] = \frac{1}{2n+1} \sum_{i=k-n}^{k+n} \left[ \frac{S[i] - \bar{S}}{\sigma_N} \right]^2 \quad (2.1)$$

ou  $\bar{S}$  et  $\sigma_N$  sont respectivement la valeur moyenne et la déviation standard du signal détecté dans un intervalle d'altitude comprise entre 20 et 22 km, ou seul le bruit est supposé présent.



**Figure 2.1.1:** Présentation des différentes étapes de traitement des données lidar : La figure de gauche montre le signal lidar corrigé de l'angle solide, ainsi que le moléculaire qui lui correspond. La figure de droite montre l'opération de seuillage effectuée sur le signal de variance.

Une valeur de seuillage  $S_{th}$  est appliquée aux profils de variances retrouvées dans le but d'identifier les couches diffusantes et leurs propriétés géométriques :

$$S_{th} = \sigma_N \cdot F \quad (2.2)$$

où  $F$  est un coefficient définissant la force du seuillage et  $\sigma_N$  correspond à l'écart type du bruit. Ce seuil utilisé doit être indépendant de l'altitude et  $F$  est calculé dans ce sens (cf. l'étude de sensibilité qui suit à propos de  $F$ ).

La valeur de seuillage est comme nous le voyons fonction de  $\sigma_N$ . D'une orbite à l'autre, d'un instrument à l'autre, les variations du rapport signal à bruit conduisent à une redéfinition du seuil utilisé.

Afin de fixer  $F$ , une analyse systématique de la sensibilité à  $S_{th}$  des fractions nuageuses retrouvées est effectuée, et ceci pour toutes les structures nuageuses observées (systèmes monocouches ou multicouches de nuages bas, moyens et haute altitude). La classification des nuages en différents niveaux d'altitude est la même que celle définie dans le protocole de ISCCP, c'est à dire grâce aux pressions limites standard de 680 hPa et 440 hPa, délimitant les classes de couches nuageuse basses (L), moyennes (M) et hautes (H) [Rossow et al, 1991].

Ce filtrage par seuillage nous amène donc à la construction d'une matrice binaire (altitude en ordonnée versus trajectoire au sol) décrivant les structures détectées et leurs limites.

Une analyse des structures de petites échelles est appliquée grâce à un filtre médian. Ce dernier permet d'analyser la vraisemblance de chacune de ces structures par l'étude des profils lidar voisins. La taille de ce filtre est définie à une échelle comparable avec celle des structures nuageuses 2D.

Nous avons choisi dans l'étude de ne considérer uniquement que l'altitude des sommets des couches nuageuses. Celles-ci sont retrouvées dans le cas des données LITE à haute résolution (de 100 m sur la verticale, correspondant aux  $(2n+1)=7$  points de la fenêtre glissante de filtrage et 700 m sur l'horizontale, i.e. sur une base tir à tir). Elles sont retrouvées sans filtrage pour le cas des données GLAS, à la résolution de 76.8 m).

Nous avons donc construit un algorithme qui comporte un certain nombre de paramètres pouvant être ajustés en fonction des caractéristiques d'observation. Les choix faits sur les valeurs de seuillage utilisées, ainsi que sur les définitions de la fenêtre glissante utilisée pour le filtrage ont une influence sur les erreurs induites par l'algorithme. Cette erreur prendra en compte les probabilités de détection et de fausse alarme.

La probabilité de l'existence d'un pixel montrant une valeur de signal supérieure au seuil de détection, dans un espace de  $M+1$  pixels d'un profil lidar à été formulée par Chazette [2001]. La probabilité de non détection  $p_n$  sera d'abord exprimée en fonction du complémentaire de la distribution normale  $F_c$  telle que :

$$p_n = 1 - F_c(x_s) \quad (2.3)$$

$$\text{avec } F_c = \frac{1}{\sqrt{2\pi}} \cdot \int_{x_s}^{\infty} \exp\left[-\frac{x^2}{2}\right] \cdot dx \quad (2.4)$$

ou  $x_s$  est la valeur centrée réduite correspondant au seuil et associée à la distribution normale du signal. La probabilité globale de fausse alarme, pour  $M$  points de mesures indépendants, pris à une autre altitude que le sommet de la structure nuageuse, est donnée par :

$$P_f = 1 - (1 - p_f)^M \quad (2.5)$$

ou  $p_f = F_c(x_B)$  est la probabilité qu'un pixel de bruit vienne au dessus du seuil et  $x_B$  est la variable centrée réduite correspondante au seuil associé avec la distribution normale centrée du bruit.  $P_f$  varie dans la zone d'altitude où les structures diffusantes sont retrouvées.

Le seuil  $X$  doit être défini de telle façon que les probabilités  $P_f$  et  $p_n$  soit minimales. Ces probabilités sont reliées et la diminution de l'un fait croître l'autre. Par conséquent, la fonction d'erreur peut être exprimée comme :

$$P_{err} = P_f + p_n - P_f \cdot p_n \quad (2.6)$$

Du fait du principe de l'algorithme, l'altitude du sommet de structure retrouvée pour une couche nuageuse donnée est influencée de façon importante par la diffusion multiple introduite par les nuages d'altitude supérieure, ce qui induit des variations lentes du signal. Néanmoins, l'effet écran des nuages bas et moyens par une couche assez épaisse de nuages hauts introduit un biais significatif qui n'est pas facile à évaluer. Quand plusieurs longueurs d'onde sont utilisées, la classification statistique précédente est effectuée de manière séparée.

## 2.2 Séparation aérosols/nuages

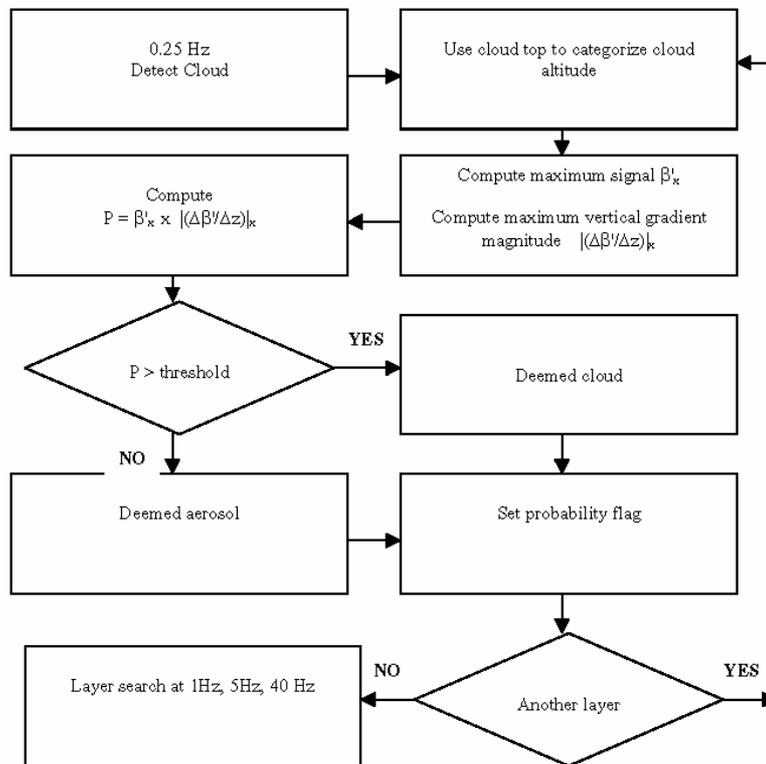
Nous choisissons de nous inspirer de la méthode utilisée par le GSFC afin de discriminer les structures nuageuses des structures aérosols dans le cadre du traitement des données lidar de GLAS. Cette méthode est fournie dans l'annexe B de cette thèse (voir la **Figure 2.2.1**). Nous présentons rapidement dans cette section comment cette méthode de discrimination fonctionne.

La technique de discrimination entre les structures nuages et aérosols repose sur l'analyse du coefficient de retrodiffusion atténuée  $b'(r)$ . La détection et la discrimination de ces structures, à partir des données GLAS, sont basées sur la connaissance que nous avons, grâce à l'observation, des différences qui les caractérisent, en terme de magnitude du signal  $b'(r)$ , et en terme de magnitude du gradient vertical de  $b'(r)$ .

A une altitude donnée, les termes  $b'(r)$  et  $|d[b'(r)]/dr|$  tendent à être plus importants dans le cas des couches nuageuses, que dans le cas des couches d'aérosols. En effet, les changements de phases de l'eau, entraînent la formation ou la dissipation des particules nuageuses.

La taille des particules, la concentration en nombre de ces particules et l'importance de l'inhomogénéité des couches nuageuses sont généralement plus importantes que celles concernant

les couches d'aérosols. Une plus complète description des bases sur laquelle repose cette discrimination entre les couches nuageuses et aérosols est donnée par Palm et al. [2002].

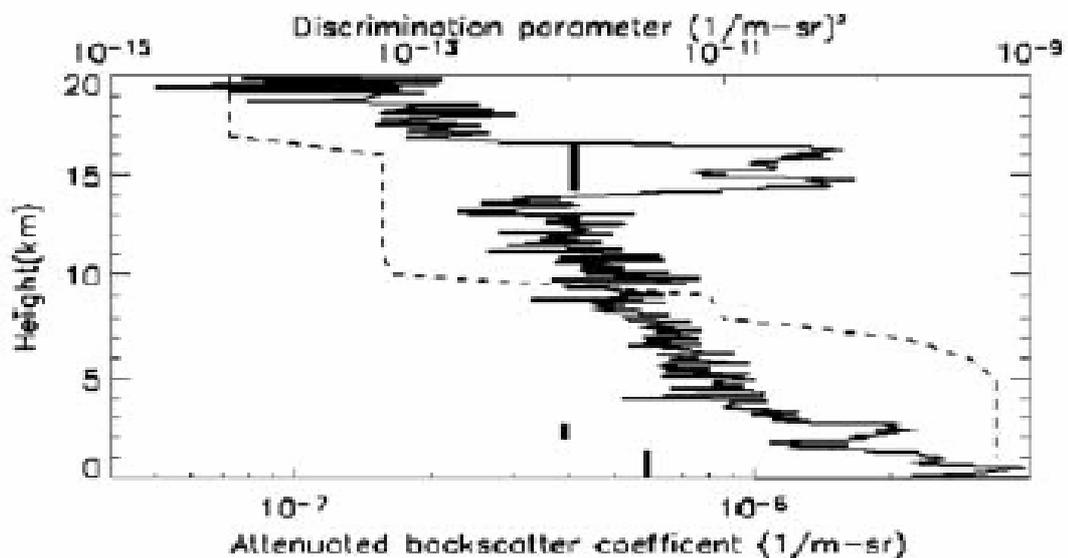


**Figure 2.2.1:** Algorithme de discrimination des couches nuageuses et aérosols (cf. ATBD GLAS, page 35, [Palm et al., 2002]).

Pour chacune des couches de particules détectées par GLAS, on détermine la valeur des termes  $b'_x(r)$  et  $|d[b'(r)]/dr|_x$ , où  $x$  représente la valeur maximale dans la couche de particules. Le produit de ces termes est calculé,  $P_l = b'_x(r) \cdot |d[b'(r)]/dr|_x$  où l'indice  $l$  correspond à la couche de particules. Le paramètre  $P_l$  est testé en fonction du seuil  $T_l$ . Si  $P_l$  est plus grand que  $T_l$ , la couche est déterminée comme étant de type nuageuse. Dans le cas contraire, cette couche est considérée comme étant de type aérosol. Si des couches de plus hautes altitudes sont présentes, une correction est appliquée à  $b'_x(r)$  et  $|d[b'(r)]/dr|_x$  avant que le produit ne soit effectué. Les pertes de transmission causées par la présence d'une couche de plus haute altitude recouvrant celle considérée sont estimées en utilisant le lidar ratio afin de calculer l'extinction en fonction de l'altitude, et en intégrant l'extinction afin d'obtenir la valeur de l'épaisseur optique et la transmission.

Le succès de cette technique est entièrement dépendant des valeurs de seuil  $T_i$  choisies. Les valeurs utilisées dans le cadre de l'extraction des données GLAS sont déterminées à partir de l'analyse des données lidar atmosphériques dérivées des mesures du lidar de haute altitude de la NASA ER-2 Cloud Lidar System [Spinhirne et al., 1982, 1996]. Dans ces études, les couches sont identifiées de manière semi-automatique par un utilisateur, comme étant des nuages ou des aérosols. Le paramètre de discrimination  $P_i$  est calculé pour ces couches de particules. A partir de ces études, une valeur de seuil  $T_i$  est choisie de manière à maximiser la probabilité d'effectuer une identification correcte de l'espèce de la couche lorsque l'on applique l'algorithme de traitement. Il a été retrouvé que les valeurs de seuil optimales de  $T_i$  sont fonction de l'altitude et par conséquent, la mise au point d'un profil vertical de  $T_i$  a été effectuée afin d'être utilisée dans l'algorithme opérationnel de GLAS. Un ajustement latitudinal de  $T_i$  est aussi nécessaire du fait que sa valeur appropriée décroisse aux régions polaires.

La **Figure 2.2.2** ([Hart et al, 2005, Fig.1 p 2]) fournit une illustration de l'application de la technique de discrimination appliquée au sein des données GLAS. On voit dans cette figure la forte dépendance du seuil de discrimination  $T_i$  en fonction de l'altitude.



**Figure 2.2.2:** Discrimination entre les couches de particules nuageuses et aérosols. La ligne solide indique  $b'(r)$ , la ligne discontinue indique le seuil de discrimination, et les barres verticales indiquent le paramètre de discrimination calculé pour chacune de ces couches. La couche supérieure est identifiée comme étant de type nuageux, alors que les deux couches de basses altitudes sont identifiées comme étant des couches d'aérosols (voir ATBD GLAS[Palm, 2002]).

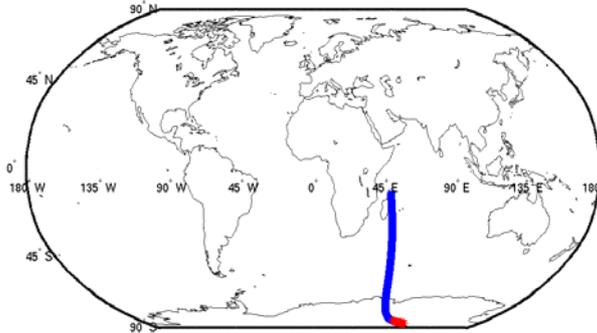
## 2.3 Etudes de cas sur les observations GLAS

### 2.3.1 Choix des cas

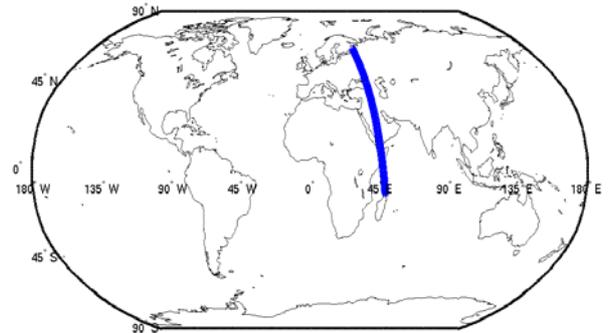
Nous avons dû dans le cadre de l'étude de cas que nous souhaitons mettre en place ici, sélectionner un ensemble de sections d'orbite dont l'étude, en tant que cas test, nous semble intéressante à faire, dans le but d'adapter et de faire évoluer les performances de notre algorithme de seuillage.

Chacun des fichiers GLA07 est composé de 10 sections, comportant chacune 36000 tirs à la résolution de 40 Hz. Nous avons retenu sur les 10 sections du fichier de données brutes **GLA07\_026\_2103\_001\_1233\_0\_01\_0001**, 4 sections d'orbites (voir la **Figure 2.3.1**). Ces 4 sections d'orbites correspondent à la journée du 13/10/2003, ce qui correspond à la campagne L2a de GLAS, c'est-à-dire utilisant le laser 2a.

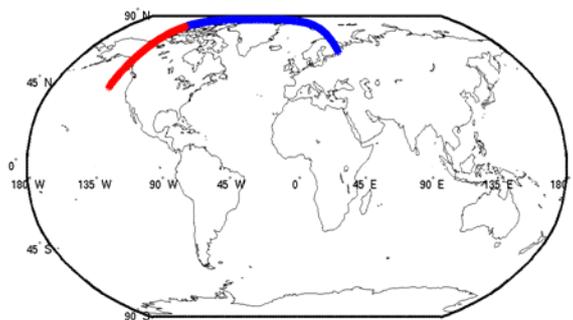
### Section 4



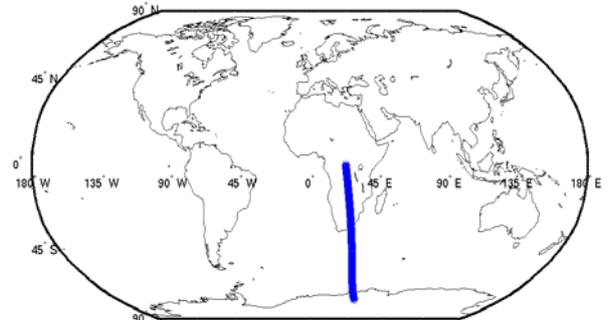
### Section 5



### Section 6



### Section 9



**Figure 2.3.1:** Localisation géographique des sections d'orbites choisie afin d'effectuer les tests de seuillage. Les horaires correspondant ainsi que les orbites sont fournis dans la table 2.1. Les données nocturnes sont tracées en couleur bleue, les diurnes en couleur rouge.

Les structures nuageuses présentes sur ces sections nous ont semblés typiques des structures nuageuses couramment rencontrées au sein de l'ensemble de la base de données GLAS.

Les horaires de chacune des sections d'orbites, et l'orbite de GLAS correspondante sont donnés dans la

**Table 2.1.**

Section	Heure de Début (GMT)	Heure de Fin (GMT)	N° de l'Orbite de GLAS
Section 4	16h39m21	16h59m31s	4077
section 5	16h59m21	17h19m40s	4077
Section 6	17h19m40	17h39m50s	4077
section 9	18h20m09	18h40m18s	4078

**Table 2.1** : Horaires (GMT) des débuts et fins de section d'orbite GLAS.

Chacune de ces sections correspond à 1200 tirs à la résolution 1 Hz, et par conséquent 6000 tirs à la résolution 5 Hz. La distance au sol séparant deux tirs lidar de GLAS à 40 Hz étant de 170 m, la distance sur l'orbite correspondante à chacune de ces sections d'orbites est de 6120 km.

Les sections d'orbites choisies sont dans leur majorité des données acquises de nuit (tracées en bleu). Les sections 4 et 6 montrent cependant la présence d'une zone de transition Jour-Nuit ainsi que des zones diurnes (tracées en rouge). Bien que notre étude soit limitée aux cas des données nocturnes, qui présentent par rapport aux données diurnes de meilleures performances en terme de rapport signal à bruit (RSB, en anglais SNR : Signal to Noise Ratio), nous avons jugé intéressant de conserver ces zones de transition. Nous pourrions ainsi juger du comportement de notre algorithme d'extraction dans ces zones, qui montrent des variations importantes, dues au changement effectué par les opérateurs du lidar dans les paramètres d'acquisition instrumentale (par exemple, le changement des filtres utilisés de jour et de nuit, le changement de l'ouverture du récepteur, ...).

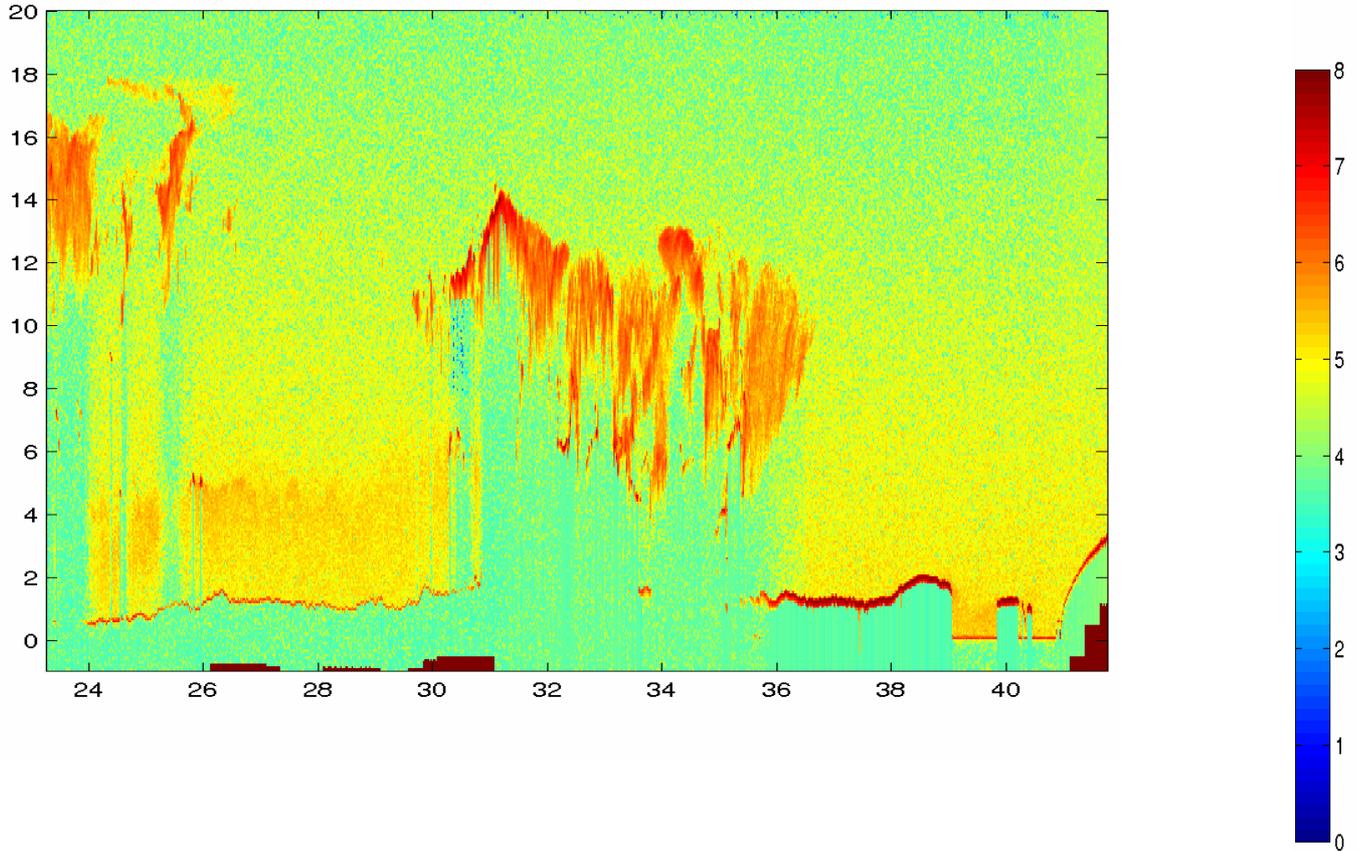
Il existe par ailleurs sur l'ensemble de la mission GLAS une évolution assez importante des performances instrumentales, et en particulier une détérioration assez notable de l'énergie d'émission des différents lasers utilisés, comme montré par [Abshire, 2005]).

Cependant, la durée séparant le début de la section 4 et la fin de la section 9 est exactement de 2 heures, ce qui reste court en regard de la durée totale de la mission. Ceci nous permet de négliger cette dépendance. Le laser 2a, à partir duquel ont été effectuées ces mesures montre en effet sur cette période une décroissance de l'énergie émise inférieure à 0.1mJ par jour, soit 5‰ de l'énergie émise.

La **Figure 2.3.2** présente le signal brut corrigé du fond de ciel, tel que fourni au sein des données GLA07, pour les sections 4, 5, 6 et 9. Pour des raisons de présentation, et du fait de la dynamique du signal, nous avons choisi d'utiliser une échelle de couleur logarithmique.

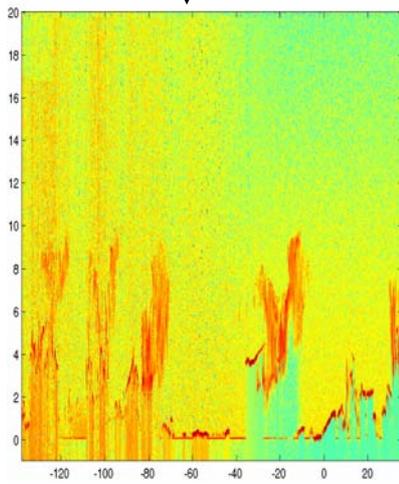
## Section 9

NUIT



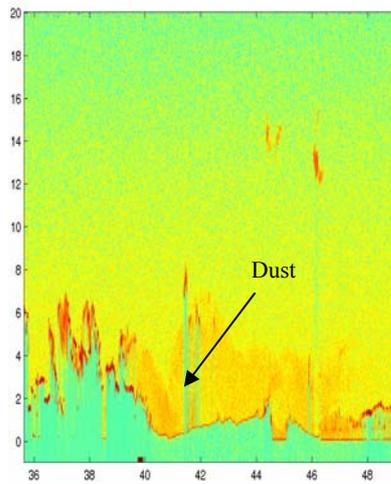
## Section 6

JOUR



## Section 5

NUIT



## Section 4

JOUR

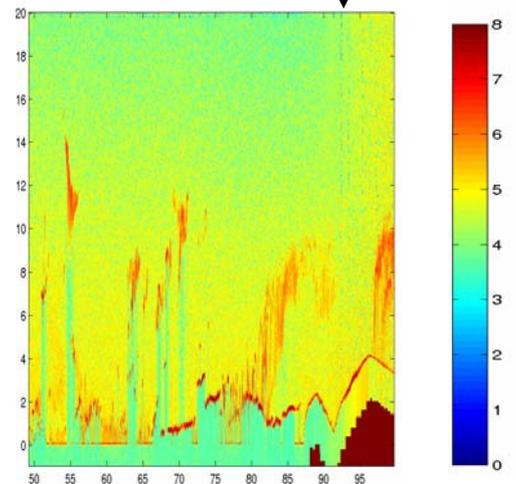


Figure 2.3.2: Logarithme du Signal brut de GLAS (corrigé du fond de ciel et de l'angle solide) sur les sections 4, 5 et 6 de l'orbite GLAS.

On remarque sur ces zones l'ensemble des nuages pouvant typiquement être retrouvés au sein des climatologies nuageuses. Nous apporterons une attention particulière lors de l'extraction de chacune de ces structures nuageuses.

On notera dans le cas des structures correspondant à des nuages semi-transparents les difficultés qui pourront être rencontrées. La structure située à l'altitude de 18 km et à 24.5°E de longitude sur la section 1, correspondant à un cirrus est en effet difficilement différenciable du bruit observé. Seule une analyse très fine des seuils à appliquer lors du seuillage, ainsi que l'utilisation de méthodes permettant l'appréhension de l'information sur les structures suivant la dimension horizontale (tirs suivant et précédant les tirs considérés) et la dimension verticale permettraient de retrouver ces structures.

On observe dans les cas des sections 4 et 6, une chute très nette de la valeur moyenne du signal au passage jour nuit. Cette chute est due à un signal de fond de ciel plus faible de nuit que de jour, lié à l'influence solaire.

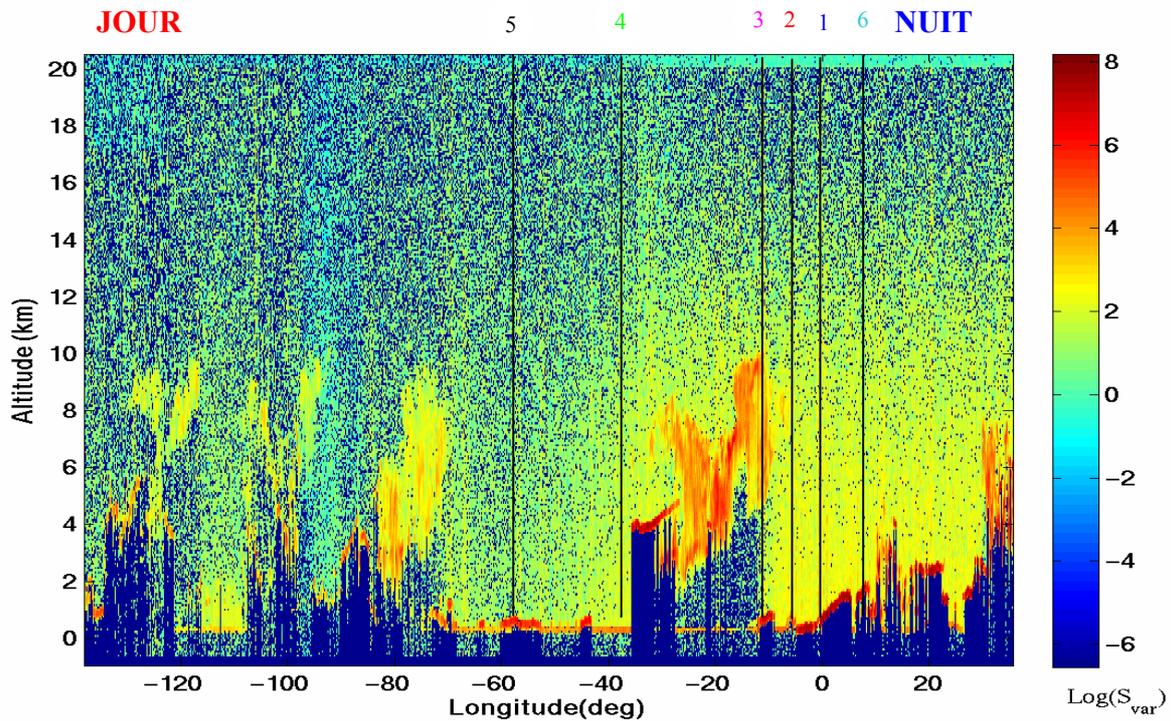
Il faudra s'attendre dans ces zones de transitions jour/nuit, où des modifications des paramètres instrumentaux de détections ont été menées afin de s'adapter aux différentes conditions de mesures, à une dégradation des performances dans les restitutions des structures nuageuses et aérosols.

On observe sur le tracé de la section 5 une zone montrant l'occurrence d'une couche d'aérosols, dans une zone géographique située à l'est de la mer noire, entre les longitudes 40°E à 48°E. L'étude des aérosols présents sur cette zone va nous permettre de tester les performances des algorithmes précédemment discutées visant à la discrimination entre les structures nuageuses et les structures aérosols.

## 2.3.2 Application

On applique par conséquent l'algorithme d'extraction sur les 4 sections GLAS présentées. Nous calculons donc dans un premier temps le signal de variance correspondant.

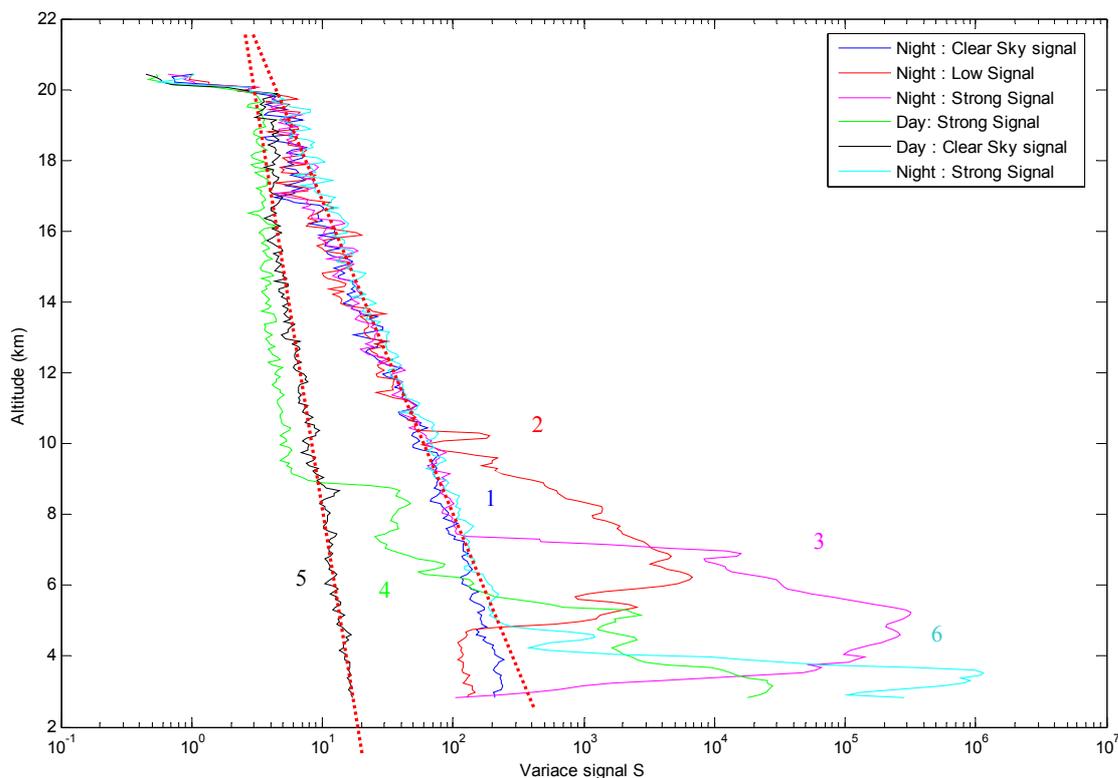
La **Figure 2.3.3** donne le signal de variance  $S_{\text{var}}$  correspondant à la section 6. On tire plusieurs éléments de réflexion suite à l'étude de cette figure :



**Figure 2.3.3:** Signal de variance  $S_{var}$  correspondant à la section 6. L'échelle de couleur est donnée en fonction du logarithme de ce signal.

On note tout d'abord un contraste important du signal lorsque nous sommes en présence de structures nuageuses, ce qui va nous permettre une extraction aisée des structures nuageuses présentes.

On observe ensuite, à partir de l'étude de cette figure, que le comportement du signal de variance semble être fonction de l'altitude considérée, et semble croître en loi de puissance à mesure que l'on se rapproche de la surface. Comme il est clairement montré dans la **Figure 2.3.4**, le biais en altitude semble particulièrement visible dans le cas des données nocturnes. Nous n'avons pas réussi dans le cadre de cette étude à déterminer la cause de ce biais.



**Figure 2.3.4:** Signal de variance  $S_{var}$  correspondant à la section 6. Les indices correspondent aux profils présentés dans la figure 1.4. Les lignes en pointillés rouges correspondent aux biais retrouvés sur les signaux.

Après avoir étudié les caractéristiques de ce biais, et vérifié sa stabilité, on propose de corriger de ce biais le signal de variance. On propose comme correction à ce biais les droites adaptées aux profils 1, 4 et 5, avec pour équation :

$$S_{fit} = 10^{\left(\frac{Z-B}{A}\right)}, \text{ avec } B = Z_0 - A \cdot \log(S_0), \quad (2.7)$$

avec  $S_0$ , la valeur du signal de variance prise à l'altitude  $Z_0$ , altitude où le signal est considéré comme exempt de structures nuageuses. Cette zone d'altitude est ici prise entre 19 et 20 km d'altitude.

La **Table 2.2** recense l'ensemble des différentes valeurs des paramètres A et B retrouvées à partir des profils 1, 4 et 5.

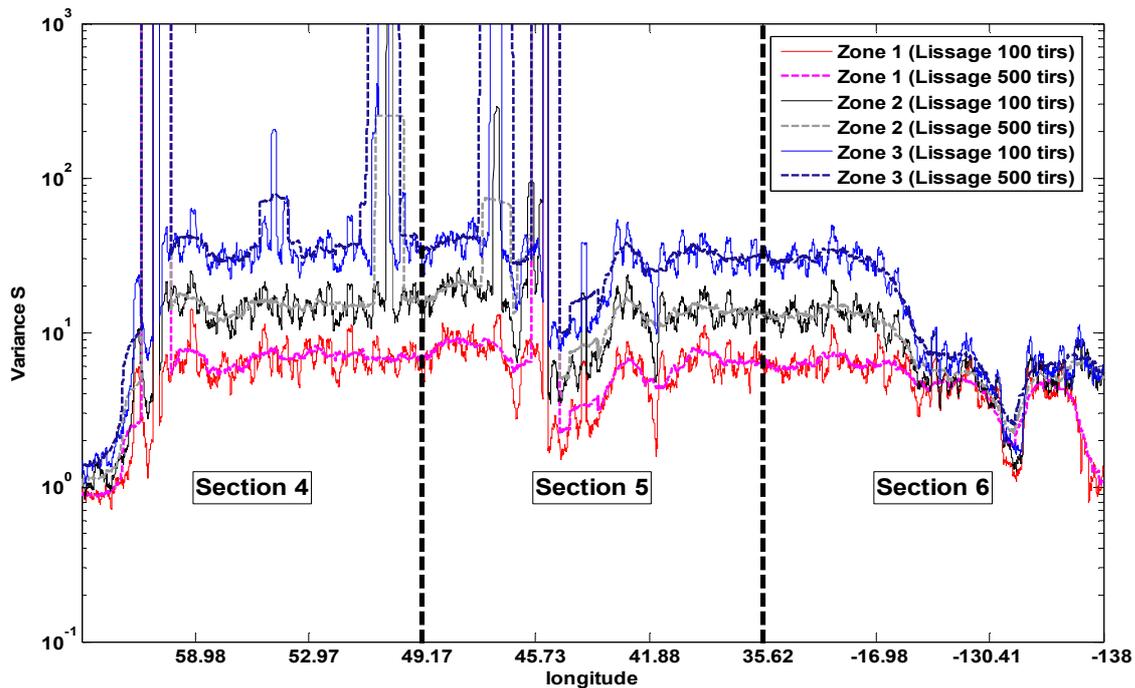
Courbes	Conditions	A	B
Courbe 1	Nuit	-8.85	25.32
Courbe 4	Jour (Dans la zone de transition Jour Nuit)	-24.61	33.44
Courbe 5	Jour	-46,47	43.28

**Table 2.2** : Paramètres A et B correspondant aux corrections de pentes appliquées au signal de Variance avant seuillage.

Après avoir effectué ces corrections de pente, nous devons maintenant prendre en compte les questions liées à la variabilité de la variance du signal de variance.

On rappelle en effet (voir méthode décrite dans la section précédente) que la valeur du seuil est proportionnelle au facteur F, conditionnant la force du seuillage. La variance de  $S_{var}$  prise à 1 Hz est jugée trop forte, et incompatible avec un mode de seuillage tir-à-tir. Son utilisation pour fixer F aurait pour conséquence de faire grandement augmenter le nombre de fausses alarmes, en faisant apparaître des structures factices fines et isolées. Il est donc nécessaire de moyennner ce signal de variance calculé à 1 Hz, grâce à une fenêtre glissante. Pour cela, nous avons effectué une étude sur la taille optimale de cette fenêtre glissante. Prendre une taille de fenêtre trop importante nous ferait perdre l'information liée à la variabilité du signal. Au contraire, choisir une fenêtre de taille trop petite nous introduirait trop de variabilité dans le seuillage.

La **Figure 2.3.5** présente pour les sections 4, 5 et 6, les courbes de la variance de  $S_{var}$  (avant correction du biais présenté dans le paragraphe précédent) en fonction des bandes d'altitudes choisies (entre 12 et 13 km (zone 1), entre 15 et 16 km (zone 2) et entre 18 et 19 km (zone 3)), et en fonction de la taille en nombre de tirs de la fenêtre de lissage utilisée (ici 100 tirs et 500 tirs). Après étude de ces courbes, nous voyons que le signal moyennné à 100 tirs montre encore une grande variabilité. Nous avons donc choisi une taille pour la fenêtre de lissage correspondant à 500 tirs, ce qui nous semble le plus approprié dans le cadre de notre algorithme de seuillage.



**Figure 2.3.5 :** Valeur de la variance du signal de variance  $S_{var}$  pour les sections 4,5 et 6. La courbe en bleu clair(bleu foncé) correspond à la variance de  $S_{var}$  calculée pour la zone d'altitude comprise entre 12 et 13 km d'altitude(Zone 1). La courbe en noir (gris) correspond à la variance de  $S_{var}$  calculée pour la zone d'altitude comprise entre 15 et 16 km d'altitude(Zone 2) et la courbe en rouge (magenta) correspond à la variance de  $S_{var}$  calculée pour la zone d'altitude comprise entre 18 et 19 km d'altitude(Zone 3), pour une fenêtre de lissage comportant 100 tirs (500 tirs).

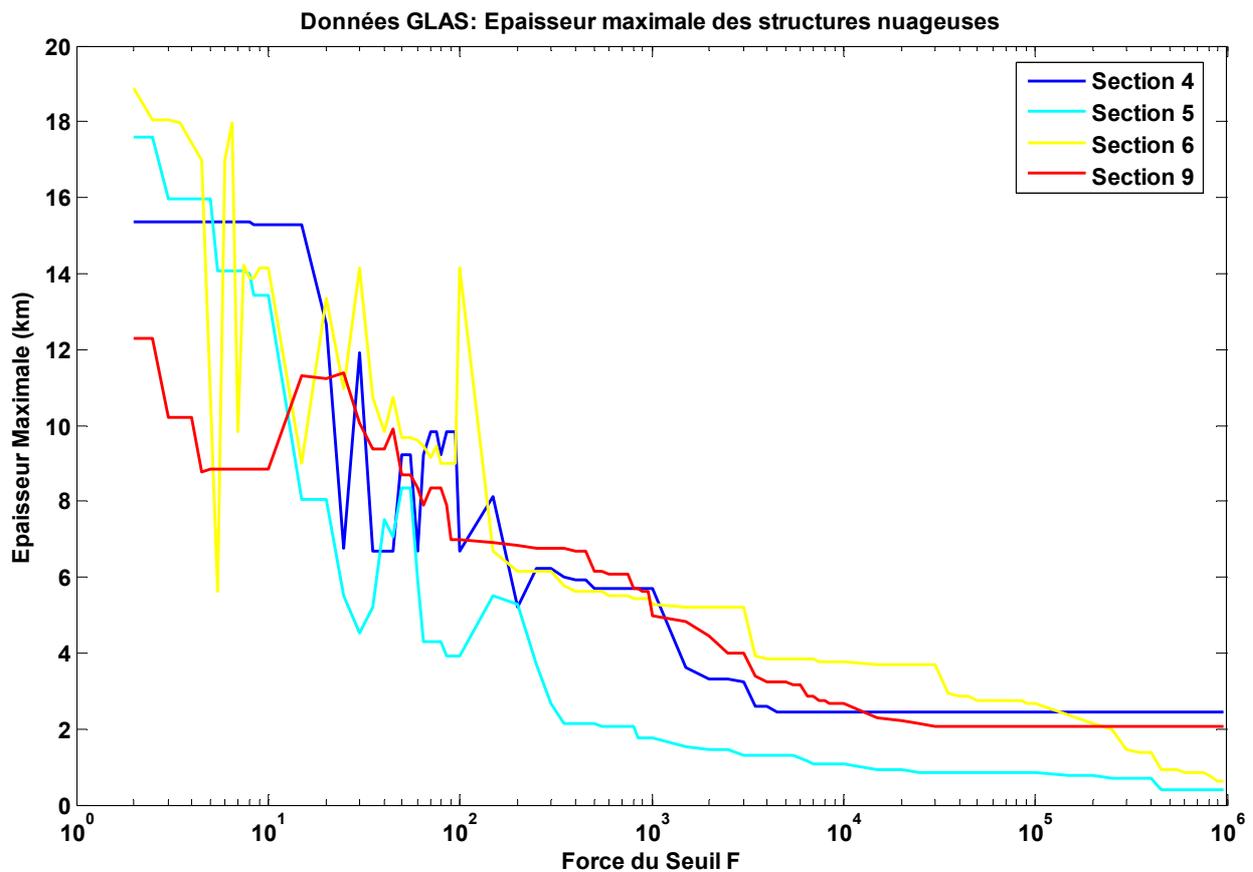
## 2.3.3 Détermination des seuils

### 2.3.3.1 Identification directe d'un seuil

Le signal de variance  $S_{var}$  ayant été calculé, et les différentes corrections ayant été appliquées, nous allons pouvoir dès lors commencer les opérations de seuillage proprement dites. Afin d'effectuer un seuillage fin du signal de variance, nous allons devoir définir quelles sont les valeurs du facteur  $F$ , caractérisant au sein de l'algorithme la force de notre seuillage que nous allons appliquer. Du fait du faible rapport signal à bruit présent sur les données GLAS, la détermination de ce facteur est un point critique à maîtriser.

Nous allons donc rechercher, sous la forme d'une étude de sensibilité, et au travers de l'étude des différents paramètres géométriques, la valeur optimale de  $F$ , permettant de minimiser la probabilité de fausse alarme, tout en maximisant la probabilité de détection des structures nuageuses.

Nous nous intéresserons dans un premier temps à la distribution de l'épaisseur des structures détectées, en fonction de la valeur du facteur  $F$  utilisée (voir la Figure 2.3.6 et la Figure 2.3.7).



**Figure 2.3.6 :** Distribution de l'épaisseur maximale des structures détectées, en fonction de la valeur du facteur  $F$  utilisée pour les 4 sections correspondant à l'étude de cas.

La figure 2.10 caractérise la distribution des épaisseurs maximales des structures détectées sur chacune des sections en fonction de  $F$ .

Pour des valeurs de force de seuillage  $F$  très faibles, l'épaisseur des structures alors déterminées sont très fortes, et bien souvent supérieures à 12 km. La valeur du seuil se trouve alors dans ces conditions bien inférieure à la valeur du signal de bruit, et par conséquent, l'ensemble du signal sur l'altitude est détecté comme une seule structure. Au contraire, pour des valeurs de force

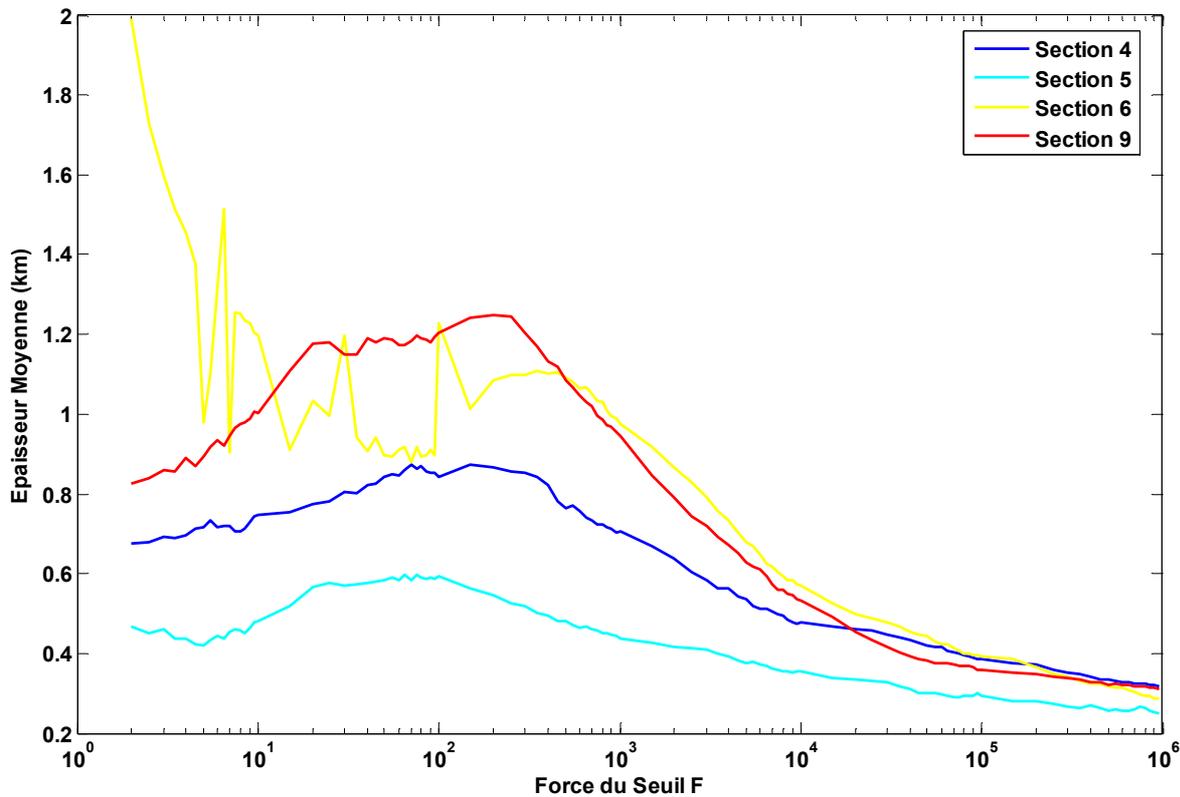
de seuil très fortes, seules les structures les plus fortes en valeur de signal « résistent » au seuillage. La valeur du seuil  $F$  qu'il nous faut maintenant déterminer est contenue entre ces deux cas extrêmes.

Il s'effectue entre ces deux extrêmes une compétition entre le nombre de structures détectées et la taille des structures détectées.

Mais nous ne percevons pas au travers de l'étude de ce paramètre d'épaisseur maximale une quelconque information nous permettant de fixer une plage de seuil possible.

Nous décidons donc de nous intéresser au paramètre d'épaisseur moyenne des structures (voir la Figure 2.3.7). Ce paramètre montre un comportement très différent du précédent. On observe en effet des valeurs plus importantes de l'épaisseur moyenne des structures, pour des valeurs du facteur  $F$  comprises entre 10 et 1000. Les valeurs maximales de chacune de ces distributions sont comprises suivant les sections regardées dans l'intervalle de valeur  $F = [100 ; 600]$ .

(Note : La section 6 a un comportement qui peut sembler marginal pour des petites valeurs de  $F$  par rapport aux sections 4, 5 et 9. A la différence des autres distributions, l'épaisseur moyenne paraît augmenter lorsque  $F$  diminue. Cet effet, a déjà été discuté dans le cadre de l'étude faite précédemment, de l'épaisseur maximale des structures, et correspond à la plage de valeurs de  $F$  inférieures au bruit, est aussi retrouvé pour les sections 4, 5 et 9, mais pour des valeurs de  $F$  inférieures à  $F=1$ , non montrées dans cette figure).



**Figure 2.3.7 :** Distribution de l'épaisseur moyenne des structures détectées, en fonction de la valeur du facteur  $F$  utilisée pour les 4 sections correspondant à l'étude de cas.

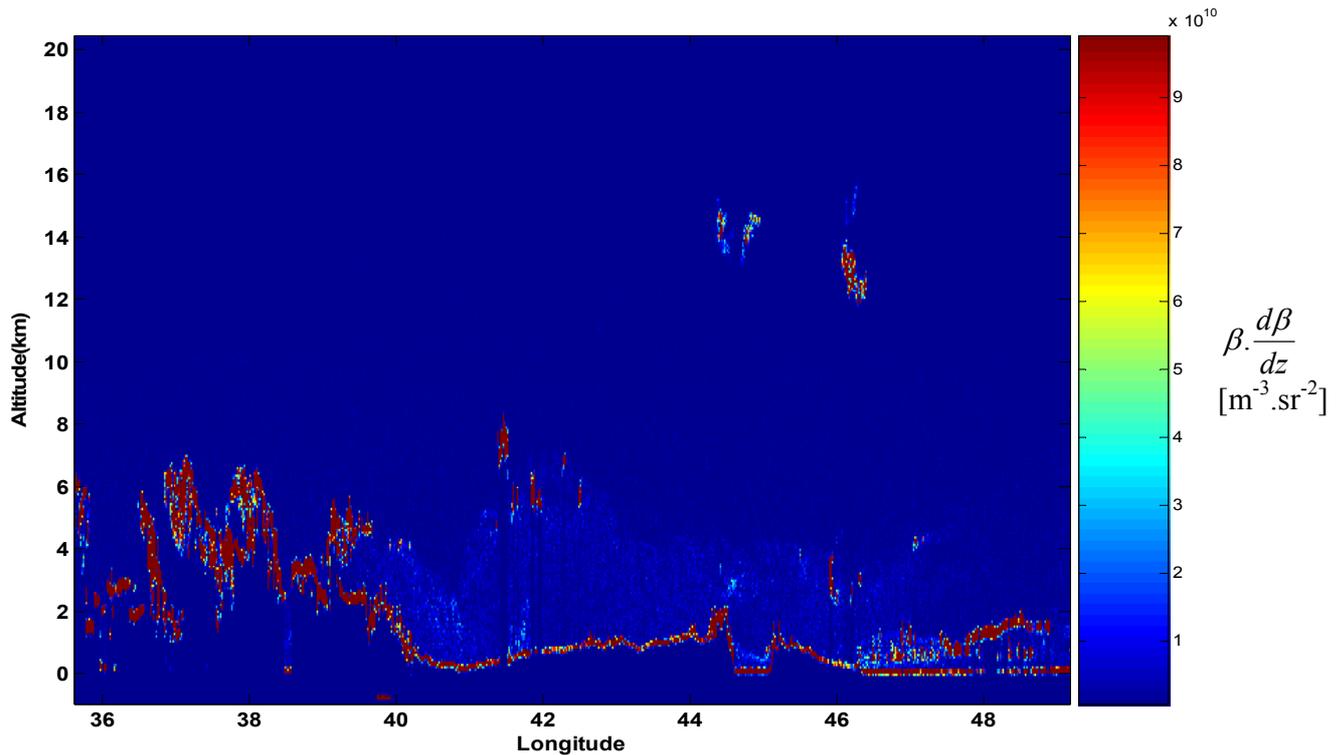
### 2.3.3.2 Séparation des nuages et des aérosols

On rappelle qu'en l'état actuel de cette étude, les distributions présentées considèrent les structures aérosols et nuageuses sans distinction. Afin de préciser la valeur optimale du facteur  $F$ , et donc par conséquent celle du seuil à appliquer, nous devons commencer à nous pencher sur cette discrimination qui doit être effectuée entre ces deux structures aux caractéristiques géométriques très différentes.

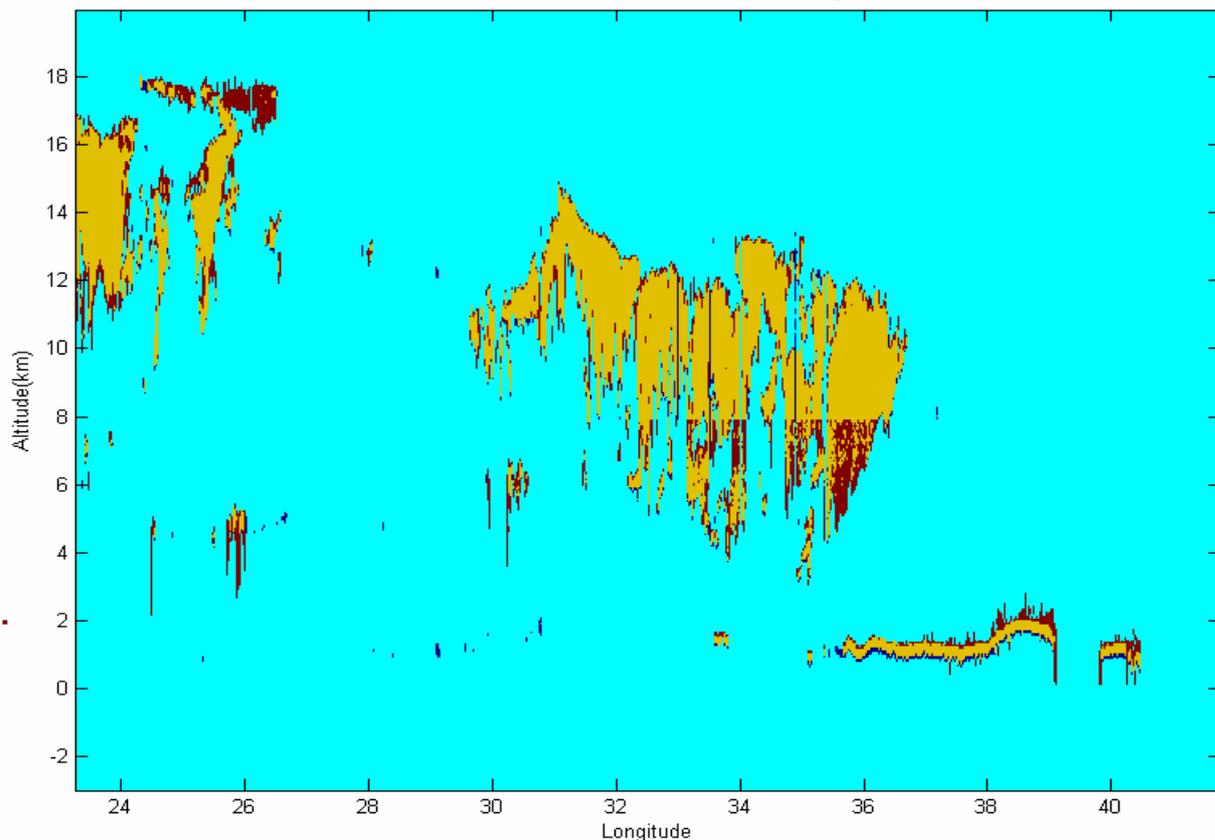
Comme décrit dans la méthode de discrimination présentée dans la section 2.2, nous identifions les structures aérosols et nuageuses entre elles grâce à l'étude du produit  $\beta \cdot d\beta$  (produit du signal de rétrodiffusion, et de la dérivée de ce signal) correspondant à chacune des structures identifiées (voir la Figure 2.3.8). Sous l'altitude de 8.5 km, les valeurs de  $\beta \cdot d\beta$  comprises sous la valeur de seuil de discrimination, notée  $S_d$ , sont identifiées comme des structures aérosols, alors que les structures supérieures à cette valeur sont identifiées comme des structures nuageuses.

La valeur  $S_d$  est prise équivalente à  $3.10^{10} \text{ m}^{-3}.\text{sr}^{-2}$  après étude du signal  $\beta.d\beta$  correspondant respectivement aux structures nuageuses et aérosols.

Afin d'éviter un effet de bord autour de l'altitude de 8.5 km (voir la Figure 2.3.9), les structures détectées comme nuages à 8.5 km, sont aussi identifiées comme tel par continuité sous cette limite arbitraire d'altitude. On peut néanmoins supposer de manière raisonnable qu'il n'y a pas d'aérosols au dessus de 8.5 km.



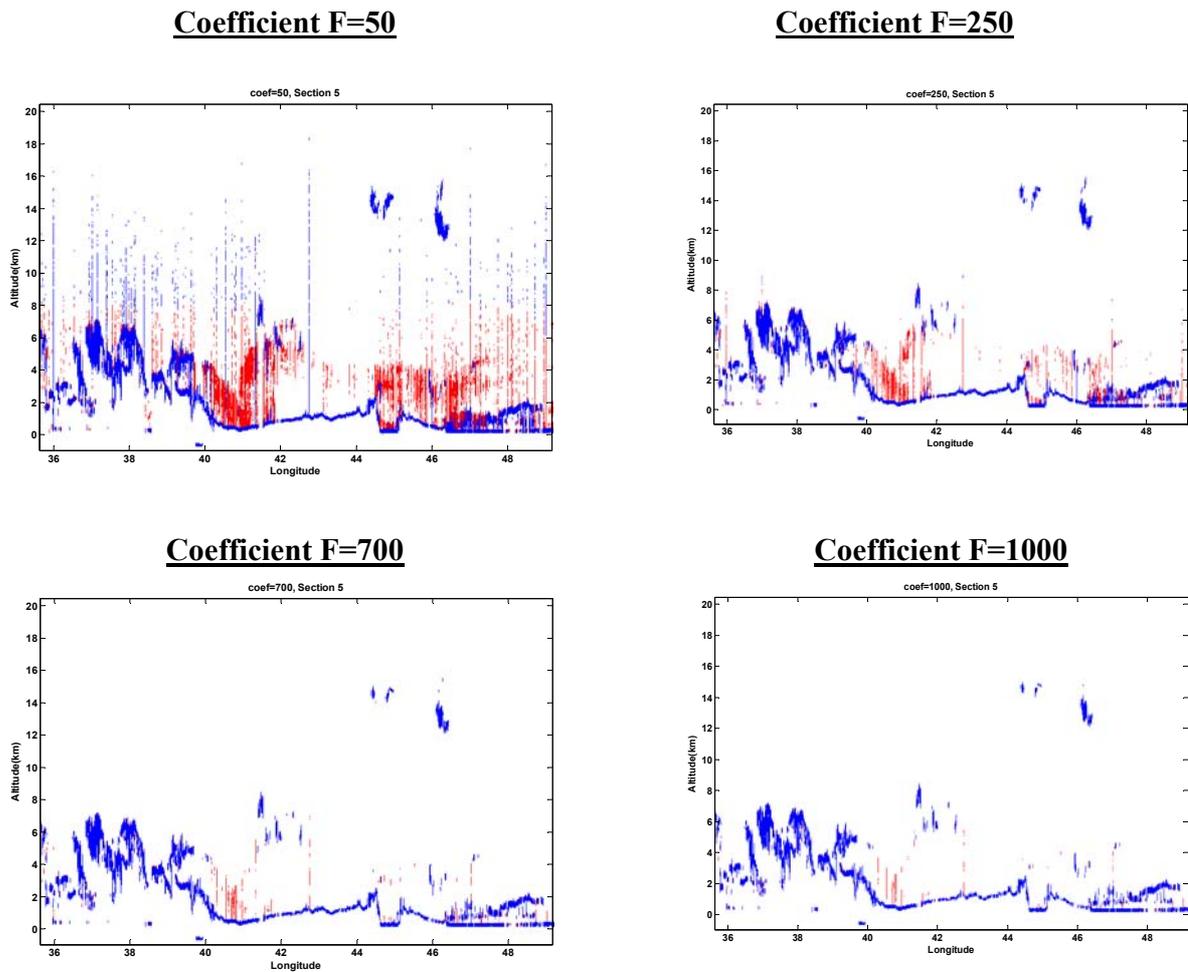
**Figure 2.3.8 :** Signal  $\beta.d \beta$  correspondant à la section 5. On note le fort contraste existant entre les valeurs de  $\beta.d \beta$  correspondant aux structures aérosols (comprises entre les longitudes 39 et 47°E), avec celui des structures nuageuses.



**Figure 2.3.9 :** Comparaison des structures nuageuses (vue en tant que surfaces), après détection par l’algorithme opérationnel et celle issue de notre algorithme. Les résultats de notre algorithme sont donnés ici avant avoir imposé la condition de continuité des structures nuageuses à l’altitude de 8.5 km. Les surfaces de couleur jaune correspondent aux surfaces détectées par les deux algorithmes, les surfaces de couleur marron à celles uniquement détectées par l’algorithme opérationnel, et les surfaces de couleur bleu foncé à celles uniquement détectées par notre algorithme.

On s’intéresse particulièrement ici à l’étude de la section 5, présentant l’occurrence de structures aérosols. Nous allons chercher à estimer quel serait le facteur  $F$  le plus approprié afin d’optimiser la probabilité de détection dans le même temps des structures aérosols et nuages, tout en minimisant la probabilité de fausse alarme (probabilité d’identifier le bruit comme une structure). La Figure 2.3.10 nous donne le résultat en sortie de l’algorithme de seuillage, correspondant à quatre valeurs de facteur  $F$ , prises dans l’intervalle de recherche déterminé précédemment. On observe que pour une valeur faible de  $F=50$ , les structures détectées sont noyées dans le bruit. Pour une valeur de seuil supérieure ( $F=250$ ), le nombre des structures faussement attribuées à une structure aérosols ou nuageuse, et correspondant au bruit, a fortement chuté. On reconnaît maintenant (tracé en rouge) la forme caractéristique d’un panache d’aérosols. Des structures

parasites isolées sont néanmoins encore détectées. Pour des valeurs de  $F$  encore supérieures ( $F=700$  et  $F=1000$ ), on élimine l'ensemble de ces structures parasites, mais ceci se fait au dépend de la détection des structures aérosols et des structures nuageuses. Après de multiples ajustements, la valeur de seuil optimale retrouvée semble se situer autour de la valeur du facteur  $F=250$ , et en tout cas comprise dans l'intervalle  $F = [100 ; 300]$ .



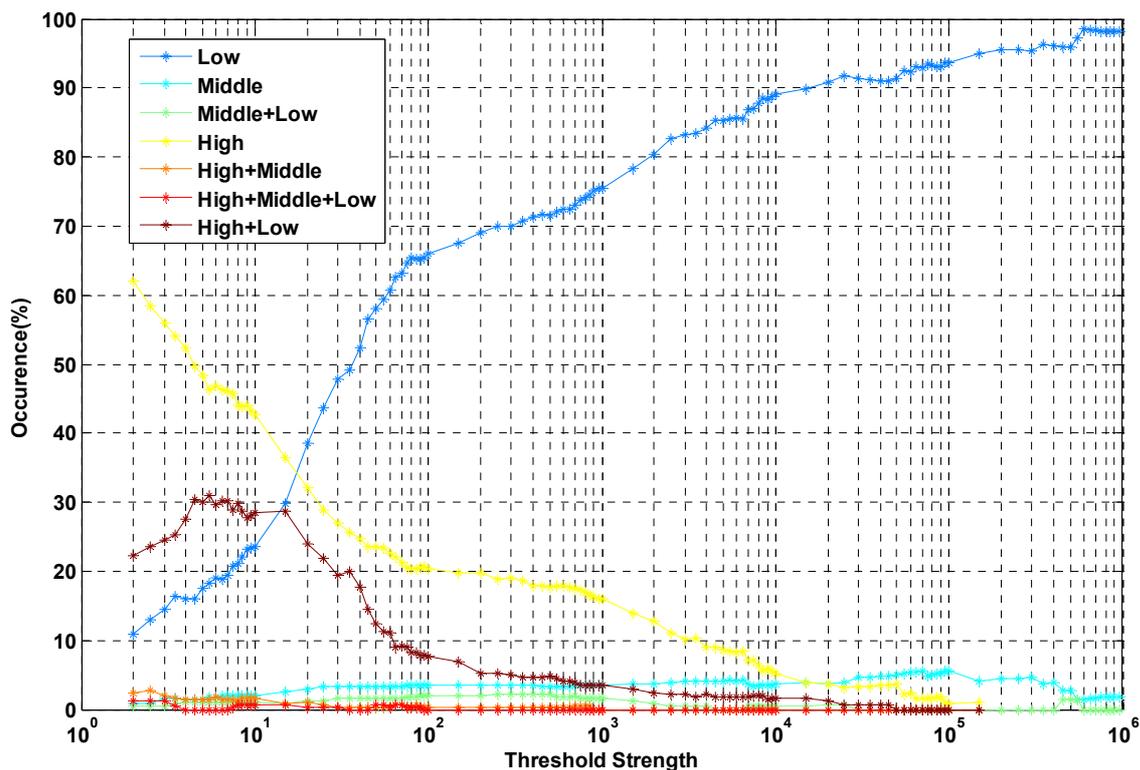
**Figure 2.3.10 :** Détection des couches aérosols (en rouge). La classe représentée en bleu est composée indistinctement des structures nuageuses et des structures correspondantes au sol.

### 2.3.3.3 Optimisation des seuils pour les nuages

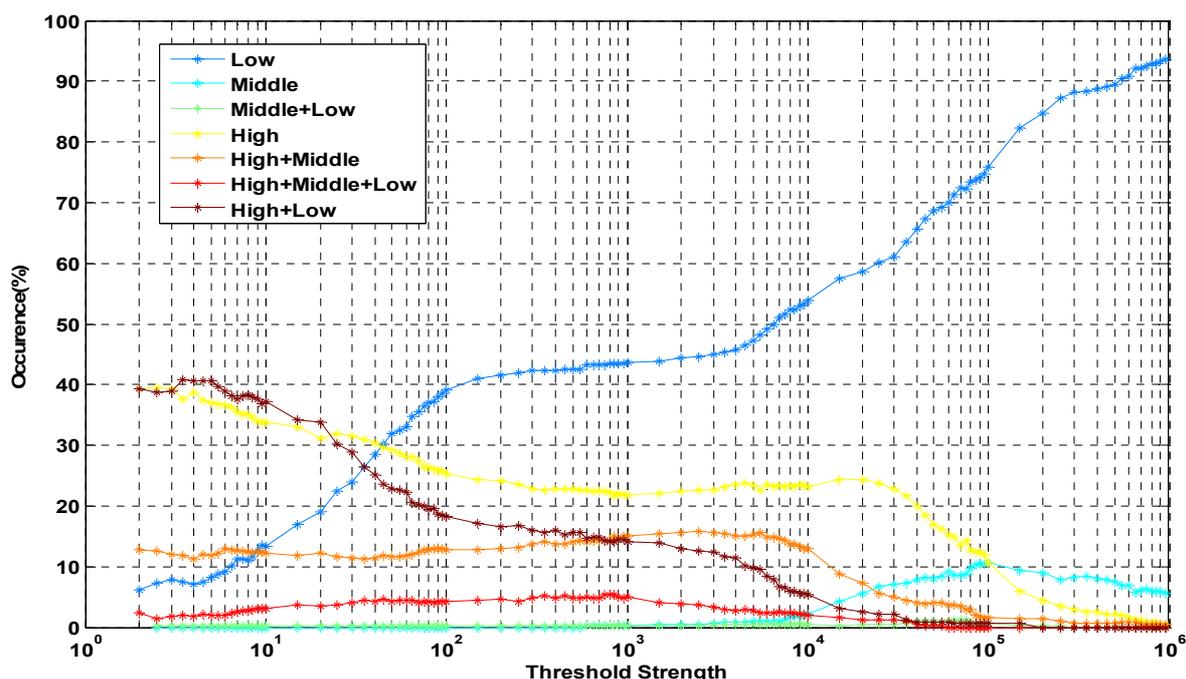
Les structures aérosols étant maintenant identifiées, nous pouvons nous affranchir de celles-ci dans nos statistiques et nous intéresser plus particulièrement à celles concernant les structures nuageuses.

Chaque tir lidar peut être classé suivant les structures nuageuses qui le composent. Ils peuvent soit être monocouches, et ne contenir alors qu'une seule structure d'altitude basse (notée B), ou de moyenne altitude (notée M), ou de haute altitude (notée H). Les tirs peuvent aussi être la combinaison de structures d'altitudes multiples, et constituer alors des tirs multicouches.

La **Figure 2.3.11** et la **Figure 2.3.12** donnent les distributions normalisées par rapport au nombre total de structure, de chacune de ces classes en fonction des valeurs de F appliquées, et ceci respectivement pour les sections 5 et 9. On note pour chacune de ces sections un palier situé sur l'intervalle des valeurs de F comprises entre F=200 et F=700. Ce palier caractérise un état d'équilibre dans la répartition de ces classes d'altitude, état d'équilibre qui semble être caractéristique de la zone de recherche des valeurs de F optimales. Il s'établit, en effet, en fonction de la valeur de seuil, une balance entre l'ensemble des classes, et la classe correspondant aux tirs monocouches comportant un nuage bas.



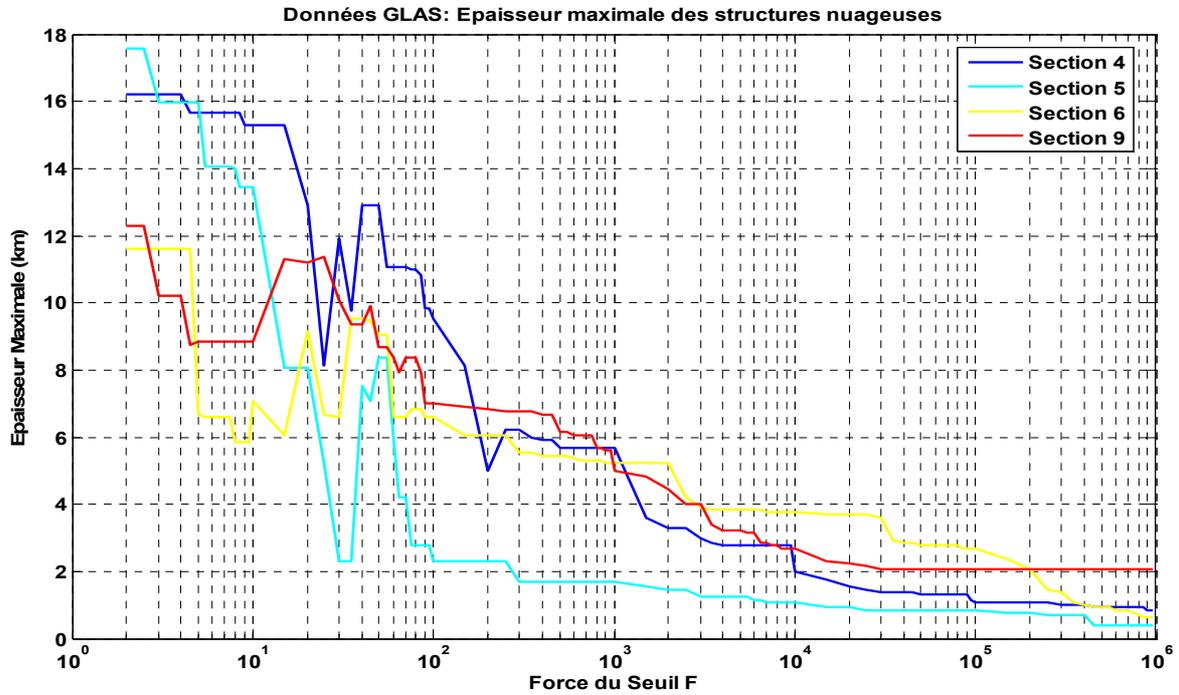
**Figure 2.3.11 :** Distributions des tirs lidar de GLAS pour la section 5, suivant les différentes classes monocouches (B, M et H) et multicouches (association de plusieurs structures classées B, M et H sur le même tir lidar). Les courbes sont données hors structures aérosols.



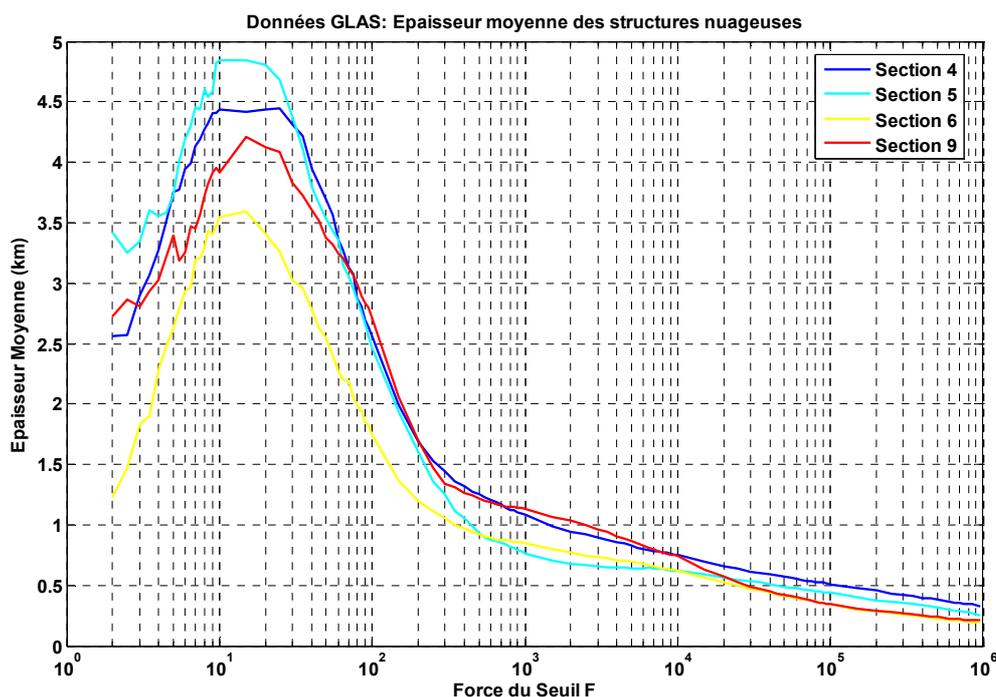
**Figure 2.3.12 :** Distributions des tirs lidar de GLAS pour la section 9, suivant les différentes classes monocouches (B, M et H) et multicouches (association de plusieurs structures classées B, M et H sur le même tir lidar). Les courbes sont données hors structures aérosols.

Nous pouvons maintenant nous pencher sur la **Figure 2.3.13** et la **Figure 2.3.14**, reprenant comme pour la **Figure 2.3.6** et la **Figure 2.3.7**, les distributions de l'épaisseur maximale et moyenne des structures détectées, mais maintenant après filtrage des structures aérosols, en fonction de la valeur du facteur F, et ceci pour les 4 sections correspondant à notre étude de cas. Comme dans le cas de la **Figure 2.3.6**, le paramètre d'épaisseur maximale (voir la **Figure 2.3.13**) ne semble pas porteur d'information sur la définition d'un seuil optimal. La **Figure 2.3.14** est par contre plus parlante sur ce sujet. L'ensemble des distributions présentées montre le même comportement en fonction du facteur de seuil F utilisé.

On remarque que les distributions sont construites selon deux modes. Un premier mode compris sur l'intervalle des valeurs  $F = [0 ; 100]$ , et montrant un optimum autour de la valeur  $F=15$ . Ce premier mode semble correspondre à la distribution des structures nuageuses en présence de bruit. Une seconde partie de la courbe a un profil plus linéaire et correspond à la distribution des épaisseurs moyennes pour des structures nuageuses hors présence de bruit. La valeur de F optimale sera celle qui séparera ces deux distributions, en minimisant ainsi la probabilité de fausse alarme.



**Figure 2.3.13 :** Distribution de l'épaisseur maximale des structures détectées, en fonction de la valeur du facteur  $F$  utilisée pour les 4 sections correspondant à l'étude de cas, après élimination des structures aérosol.

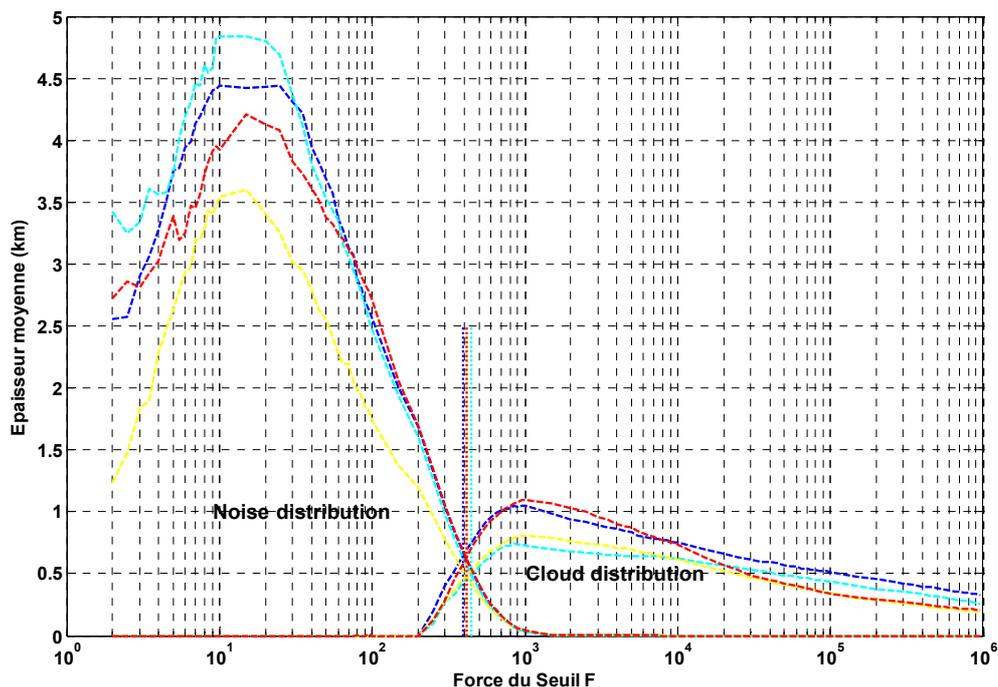


**Figure 2.3.14 :** Distribution de l'épaisseur moyenne des structures détectées, en fonction de la valeur du facteur  $F$  utilisée pour les 4 sections correspondant à l'étude de cas, après élimination des structures aérosol.

A partir des distributions précédentes, nous pouvons chercher à estimer la valeur de force de seuil  $F$  telle que la probabilité de fausse alarme soit minimum, tout en maximisant la probabilité de bonne détection (voir la Figure 2.3.15). Nous prolongeons la distribution correspondant à celle du bruit, et faisons la déduction en conséquence de la distribution en épaisseur des nuages.

Nous cherchons alors, pour chacune des sections, le point pour lequel la distribution correspondant au mode bruité, et celle correspondant à la distribution nuageuse sont égales. Nous trouvons pour les sections 4, 5, 6 et 9, respectivement les valeurs de force de seuil  $F_4=395$ ,  $F_5=441$ ,  $F_6=413$  et  $F_9=411$ .

En utilisant ces valeurs de force de seuil, nous dégraderions trop fortement la distribution nuageuse. Un seuil plus faible paraît donc préférable. Nous choisissons par conséquent de conserver l'ensemble de la distribution nuageuse, ce qui est faisable en choisissant une valeur de force de seuil égale à  $F=250$ .



**Figure 2.3.15 :** Distribution de l'épaisseur moyenne des structures nuageuses et du bruit, en fonction de la valeur du facteur  $F$  utilisé pour les 4 sections, et après élimination des structures aérosol.

## 2.3.4 Résultats sur les cas tests et intercomparaison à l'algorithme opérationnel

### 2.3.4.1 Corrélation des sommets de structures

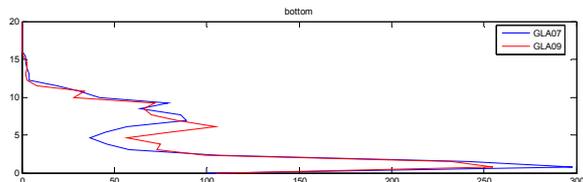
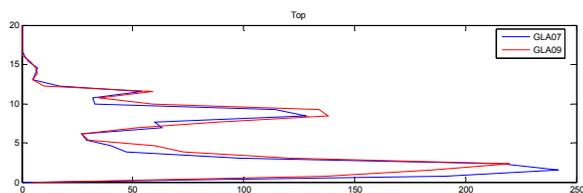
Dans la partie précédente, nous avons mis en place un algorithme qui va nous permettre d'effectuer le seuillage des données lidar de la mission GLAS, et d'effectuer ainsi une analyse statistique à l'échelle globale de la distribution spatiale des structures nuageuses.

En vue de valider notre algorithme d'extraction, nous avons choisi d'appliquer notre algorithme de seuillage au cas des données GLAS, et de comparer les résultats alors obtenus avec ceux fournis grâce à l'algorithme opérationnel de GLAS.

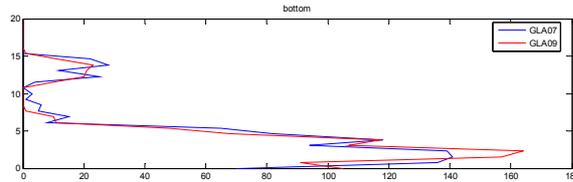
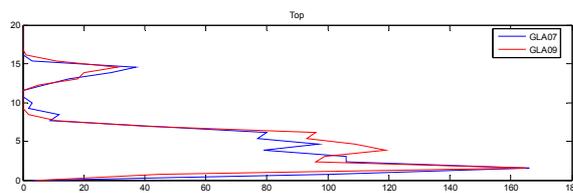
Nous avons ainsi souhaité par cette démarche montrer la souplesse et l'évolutivité de notre algorithme, et en particulier montrer la robustesse de ce dernier dans des conditions de SNR faible, telle que présente dans le cas de la mission GLAS, ou comme c'est le cas aussi de la mission CALIPSO. Les performances annoncées de ce capteur, ainsi que les premières données enregistrées et récemment rendues publiques, nous amènent à penser en effet que ces données seront caractérisées par un SNR équivalent à celui retrouvé grâce à l'instrument GLAS.

L'algorithme de traitement utilisé par le GSFC (Goddard Space Flight Center) afin de produire le produit opérationnel de GLAS est explicité dans l'ATBD de GLAS [Palm, 2001], dont un extrait est fourni en annexe de ce document de thèse.

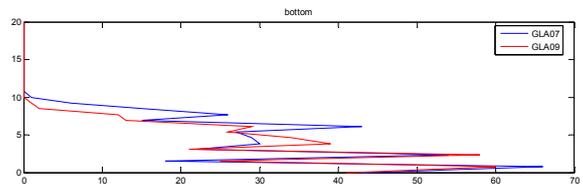
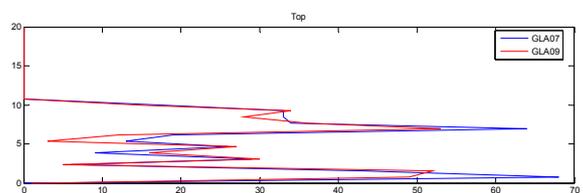
Après avoir choisi la force de seuil  $F$  équivalente à 250 comme définie dans la précédente section, nous produisons à partir des données brutes GLA07 et ceci pour chacune des sections, les distributions en altitudes des sommets de structures nuageuses (données GLA07, tracées en bleu dans la **Figure 2.3.16**). Ces fonctions de densité de probabilité (notée FDP) (en anglais PDF, pour Probability Density Function) sont directement comparées avec celles retrouvées grâce à l'algorithme opérationnel (données GLA09, tracées en rouge dans la **Figure 2.3.16**).



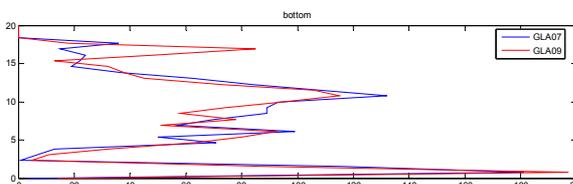
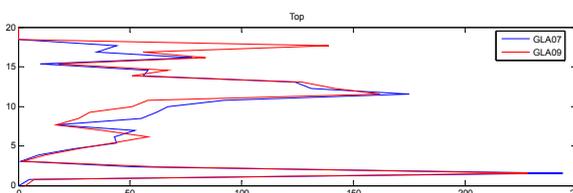
Section 4



Section 5



Section 6



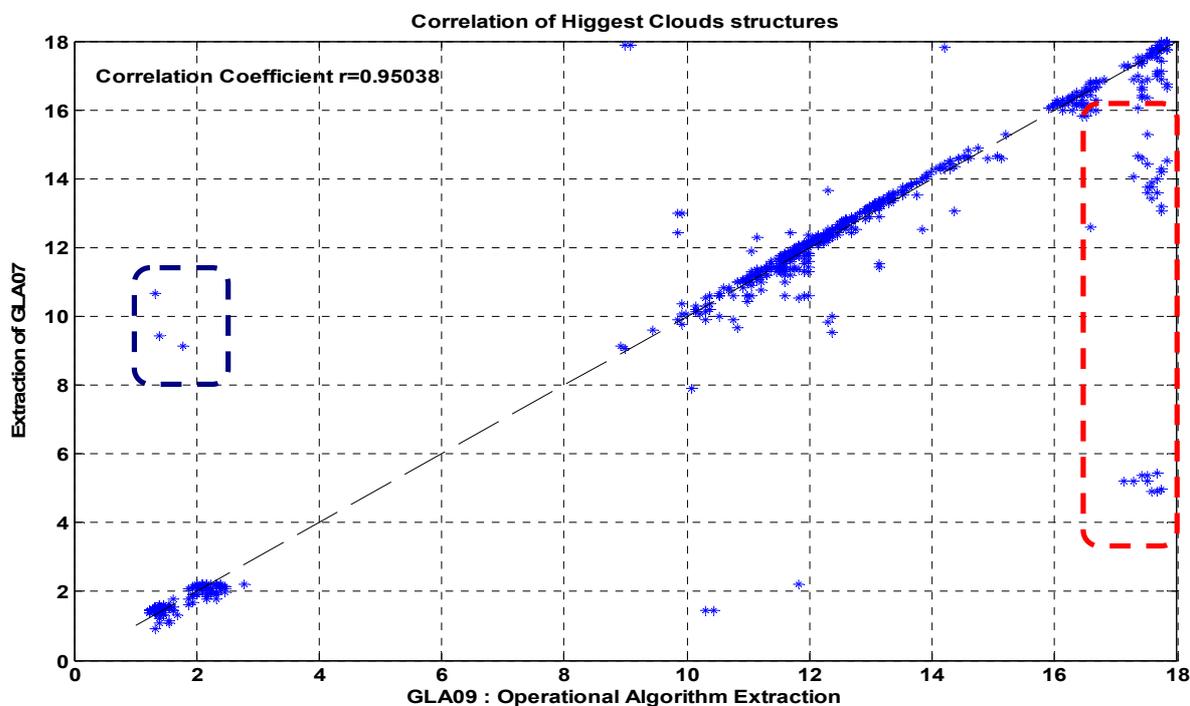
Section 9

**Figure 2.3.16 :** PDF des sommets et des bases de structures nuageuses retrouvées grâce à l'extraction par notre algorithme des données brutes GLA07. Les PDF correspondantes retrouvées par l'algorithme opérationnel sont fournies en comparaison.

Les PDF des sommets de structures correspondent dans l'ensemble des sections avec les sorties de l'algorithme opérationnel, ce qui semble conforter la validité de notre algorithme. Quelques différences sont retrouvées (voir la **Figure 2.3.16**, section 9) pour les structures de hautes altitudes optiquement faibles, et dont la détection est critique, car leur signal est de même intensité que le signal correspondant au bruit.

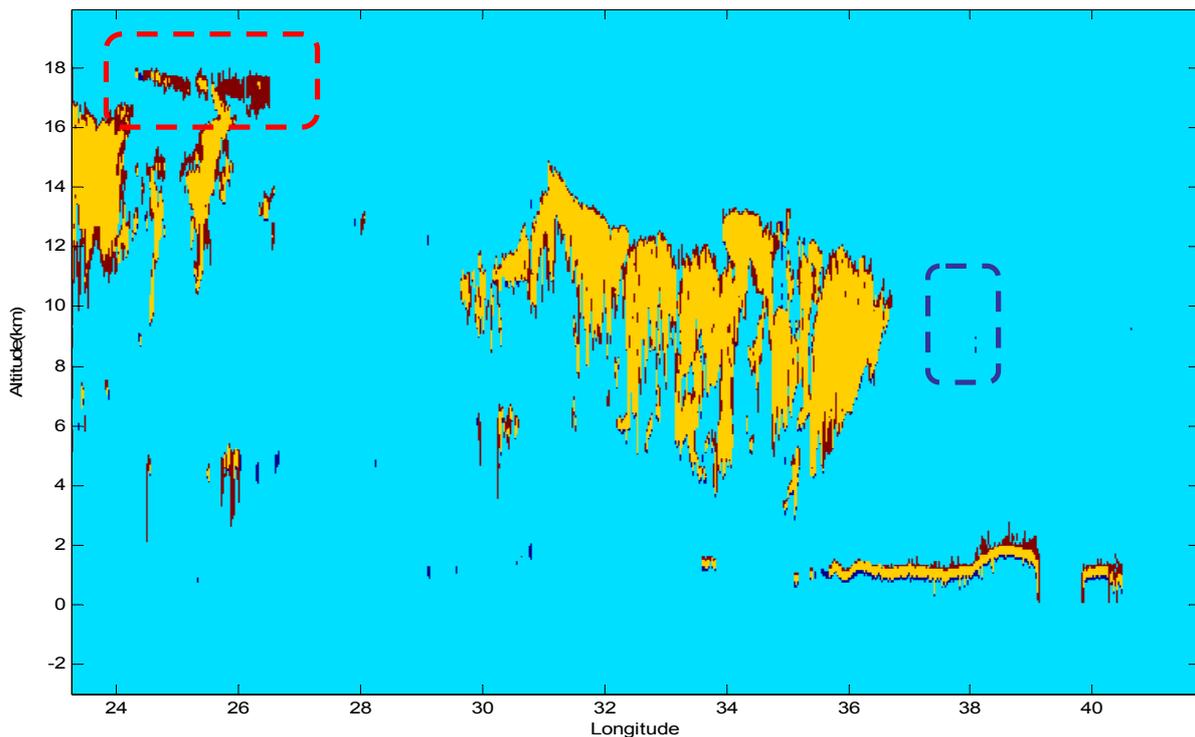
Comme le montre la **Figure 2.3.17**, donnée pour le cas de la section 9, de très bonnes corrélations sont néanmoins retrouvées (Les points situés à gauche de la droite de corrélation correspondent à des sommets de structures non vus par l'extraction de l'algorithme opérationnel,

alors que ceux situés à droite correspondent à des sommets de structures non vus par notre algorithme).



**Figure 2.3.17 :** *Corrélation des sommets de structures, retrouvés grâce à l'utilisation des deux algorithmes d'extraction des données GLAS, sur la section 9.*

La valeur du facteur de corrélation oscille autour de  $r=0.95$ . On observe que certaines structures ne sont pas détectées par notre algorithme. Ces sommets correspondent dans le cas de la section 9 aux sommets du cirrus compris entre les longitudes 24 et 25°E, à 18 km d'altitude (voir encadré en tirets rouges de la **Figure 2.3.17**, et de la **Figure 2.3.18**). Ce cirrus, partiellement détecté, explique les différences dans les PDF observées à ces altitudes (voir la **Figure 2.3.16**, Section 9). Notre extraction fait ressortir aussi des structures non détectées par l'algorithme opérationnel de GLAS (voir encadré en tirets bleus de la **Figure 2.3.17**, et de la **Figure 2.3.18**).



**Figure 2.3.18** : Comparaison des structures nuageuses (vue en tant que surfaces), après détection par l'algorithme opérationnel et celle issue de notre algorithme dans sa version finalisée ( $F=250$ ). Les surfaces de couleur jaune correspondent aux surfaces détectées par les deux algorithmes, les surfaces de couleur marron à celles uniquement détectées par l'algorithme opérationnel, et les surfaces de couleur bleu foncé à celles uniquement détectées par notre algorithme.

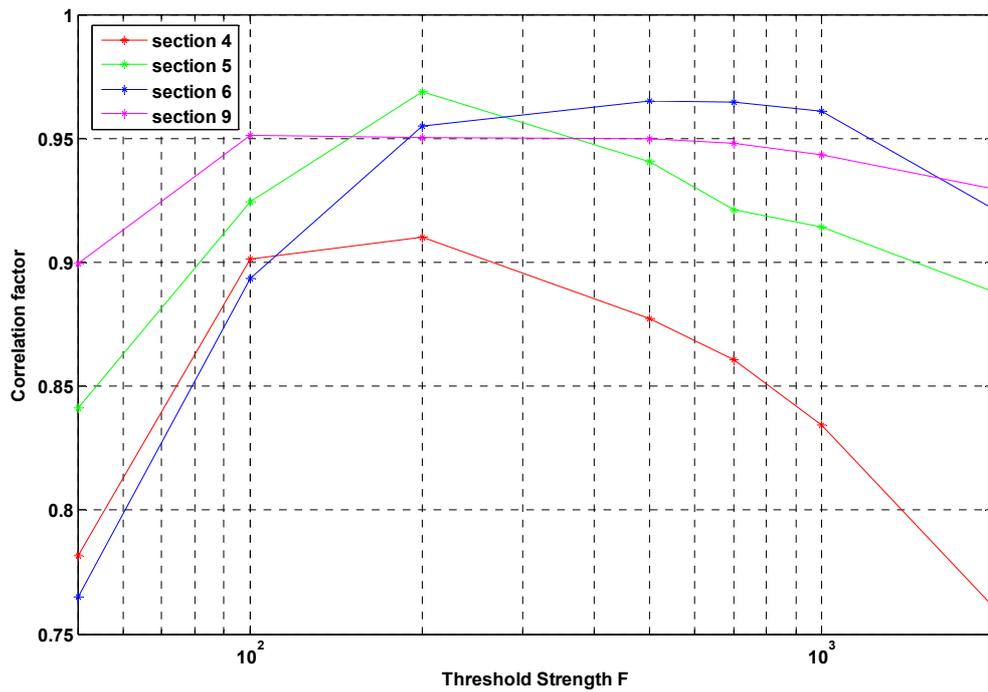
Nous profitons de cette analyse pour valider la définition de la valeur du facteur de force de seuillage  $F=250$  effectuée dans la précédente section.

Nous cherchons donc à voir comment évolue la corrélation des résultats obtenus par l'algorithme opérationnel et le nôtre, en fonction de la force du seuillage  $F$ . Nous devrions ainsi retrouver un optimum. Nous confrontons donc entre elles les altitudes des sommets de structures nuageuses les plus hautes sur chacun des tirs lidar.

La corrélation est estimée grâce au coefficient de corrélation de Pearson  $r$ , dont la formule pour deux ensembles  $X$  et  $Y$  est :

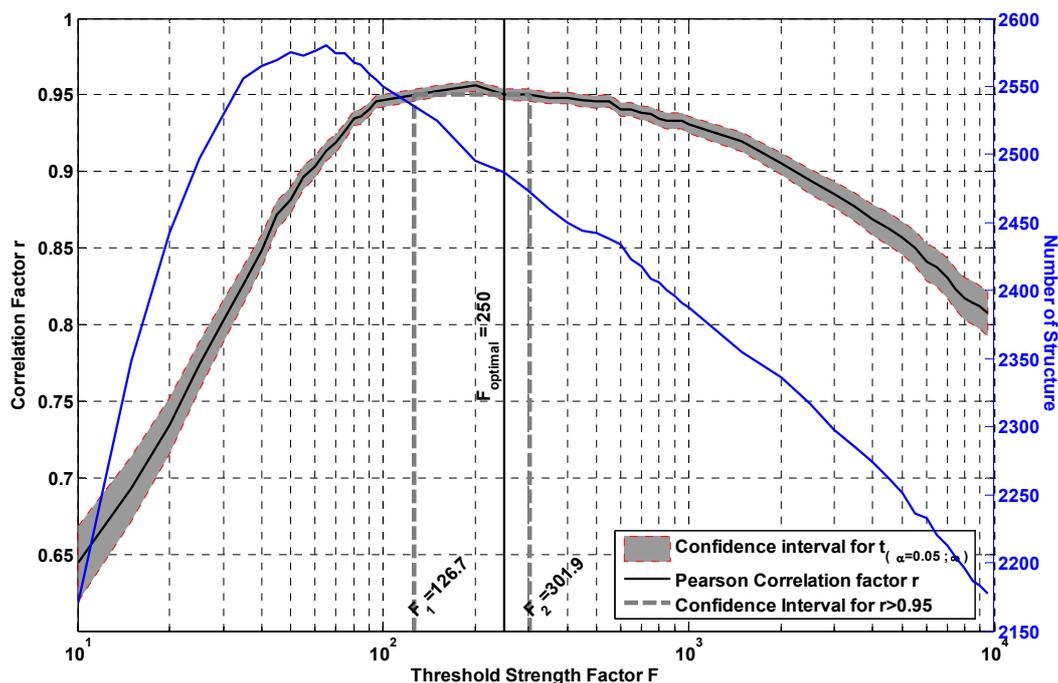
$$r_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N \cdot \sigma_X \cdot \sigma_Y} \quad (2.8)$$

La **Figure 2.3.19** donne l'évolution du coefficient de corrélation  $r$  en fonction de  $F$ , et pour chacune des sections 4, 5, 6 et 9.



**Figure 2.3.19 :** Coefficient de corrélation retrouvé pour chacune des sections, en fonction du facteur de force de seuillage  $F$ .

On observe toutes sections confondues un optimum pour des valeurs proches de  $F=250$ . La **Figure 2.3.20**, correspondant au calcul du coefficient de corrélation  $r$  pour l'ensemble des sections, nous donne une valeur optimale de 250, avec un intervalle de confiance où  $r > 0.95$  compris entre  $F=126$  et  $F=301$ .



**Figure 2.3.20 :** Coefficient de corrélation de Pearson, calculé à partir des sorties des deux algorithmes, pour l'ensemble des 4 sections, en fonction du facteur de force de seuillage  $F$ . L'enveloppe grise donne l'intervalle de confiance sur le facteur de corrélation de Pearson  $r$ , calculé à partir de la table de student, avec  $\alpha=0.05$ , et pour un nombre de points  $> 400$ ). Le nombre de points de corrélation est donné dans la courbe bleue. L'intervalle de confiance est donné pour les valeurs du coefficient de corrélation supérieures à 0.95.

Cette étude confirme la valeur de  $F$  choisie précédemment comme équivalente à 250. La très forte corrélation obtenue entre les résultats de l'algorithme opérationnel et le nôtre, confirme la qualité de notre algorithme.

## 2.3.4.2 Comparaison des structures nuageuses

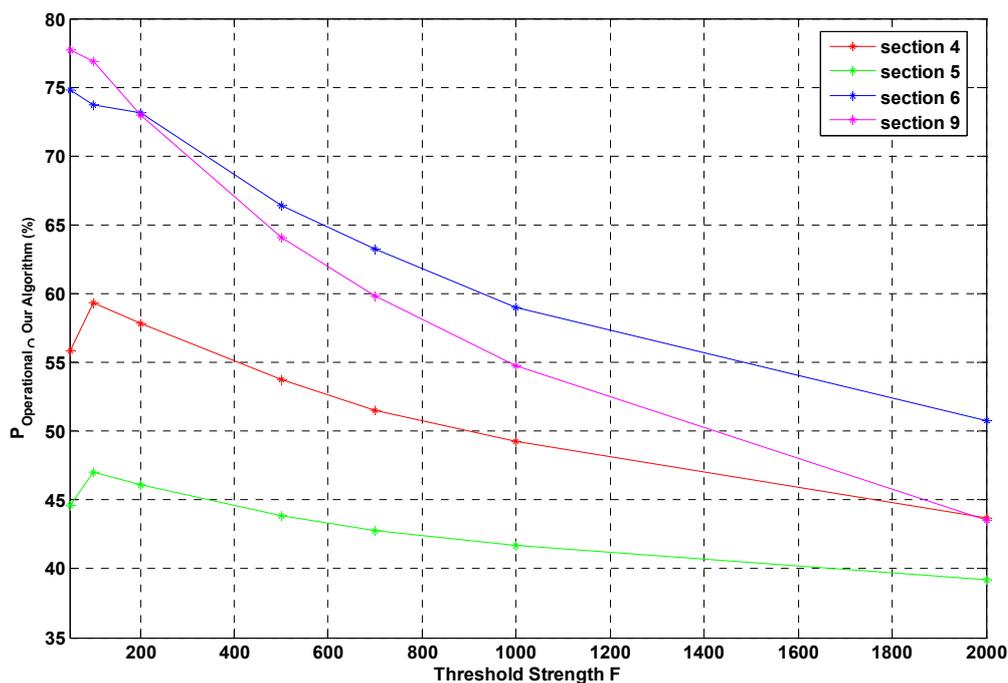
### 2D

A partir des matrices de détections telles que fournies dans la **Figure 2.3.18**, nous pouvons calculer les corrélations en terme de surface commune des structures (c'est-à-dire en nombre de pixels), et non plus en terme de sommets de structures. On donne ainsi un nouvel indice sur les performances des algorithmes. La **Figure 2.3.21**, donne pour chacune des 4 sections test, la proportion des surfaces de détection commune, ramenée à l'ensemble des observations, pour les deux algorithmes et en fonction de la force de seuillage  $F$ . Les mêmes courbes sont fournies aussi

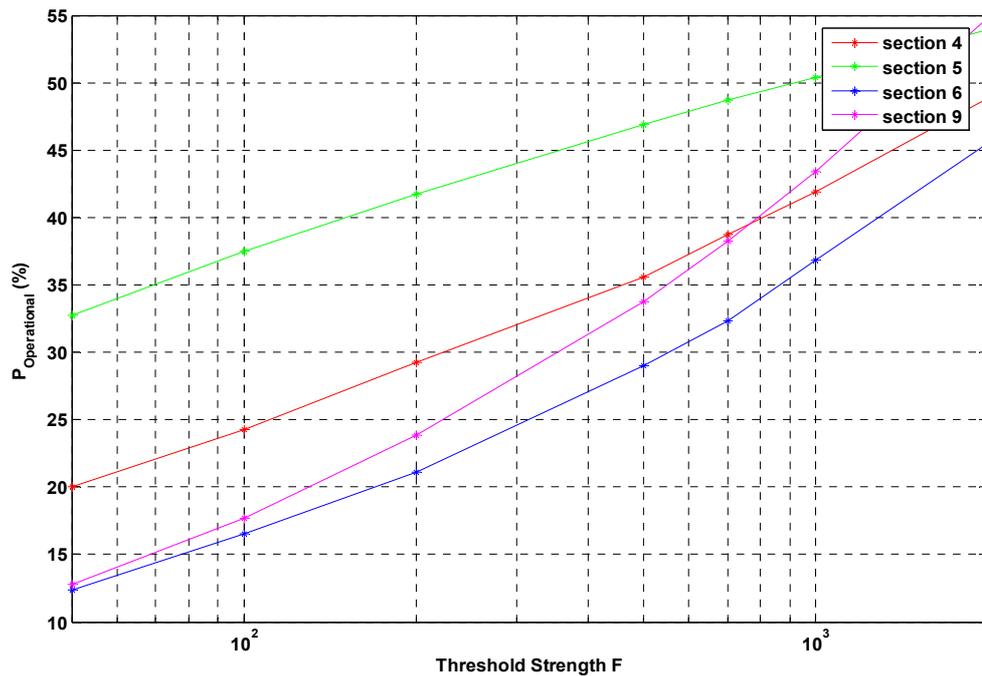
pour les détections de surfaces faites uniquement par l'algorithme opérationnel (voir la **Figure 2.3.22**), ou notre algorithme (voir la **Figure 2.3.23**).

Nous voyons qu'en terme de surface, les résultats paraissent moins bons que ceux précédemment présentés grâce au facteur de corrélation  $r$ . On observe en effet pour  $F=250$ , une probabilité de détection commune comprise entre 46 et 73 % (voir la **Figure 2.3.21**), une probabilité de détection uniquement faite par l'algorithme opérationnel comprise entre 16 et 37% (voir la **Figure 2.3.22**), et une probabilité de détection uniquement faite par notre algorithme comprise entre 5.5 et 17 % (voir la **Figure 2.3.23**).

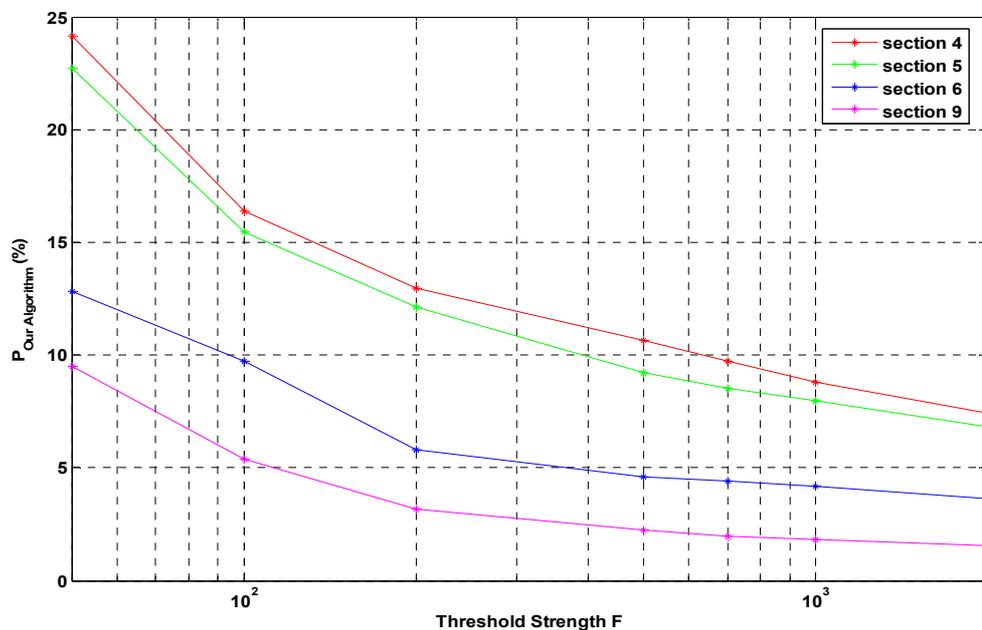
Une des principales causes d'erreurs semble être liée à une détection des bords de structures moins performante que dans le cas de l'algorithme opérationnel de GLAS. Ce constat est particulièrement vérifié dans le cas de la détection des bases de structures nuageuses. On observe de plus dans certain cas une détection défailante sur les nuages semi transparents de hautes altitudes tels que les cirrus.



**Figure 2.3.21 :** Proportion du nombre de pixels nuageux détectés de façon commune par les deux algorithmes, en fonction de la force de seuillage  $F$ . Ce nombre de pixels est ramené à l'ensemble des observations. Les proportions sont données pour chacune des sections.



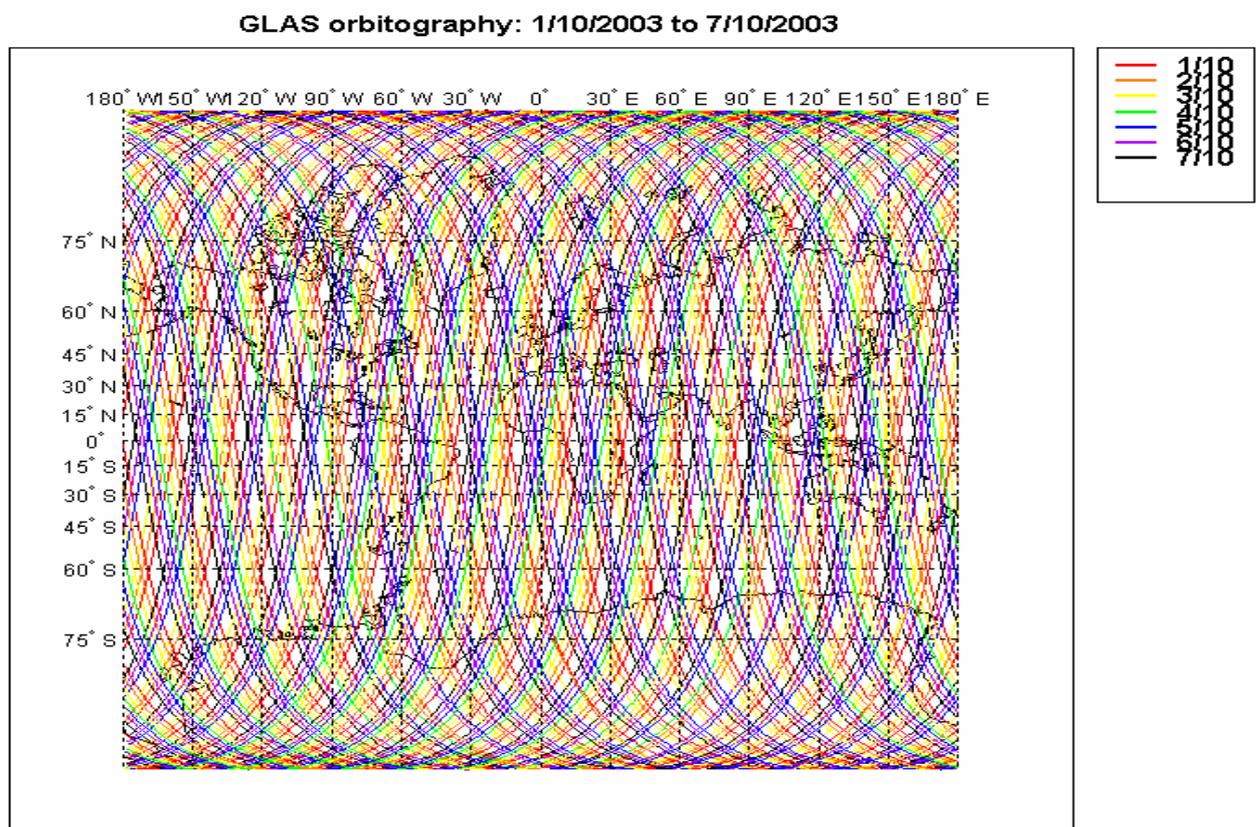
**Figure 2.3.22 :** Proportion du nombre de pixels nuageux détectés par l'algorithme opérationnel de GLAS uniquement, en fonction de la force de seuillage  $F$ . Ce nombre de pixels est ramené à l'ensemble des observations. Les proportions sont données pour chacune des sections.



**Figure 2.3.23 :** Proportion du nombre de pixels nuageux détectés par notre algorithme uniquement, fonction de la force de seuillage  $F$ . Ce nombre de pixels est ramené à l'ensemble des observations. Les proportions sont données pour chacune des sections.

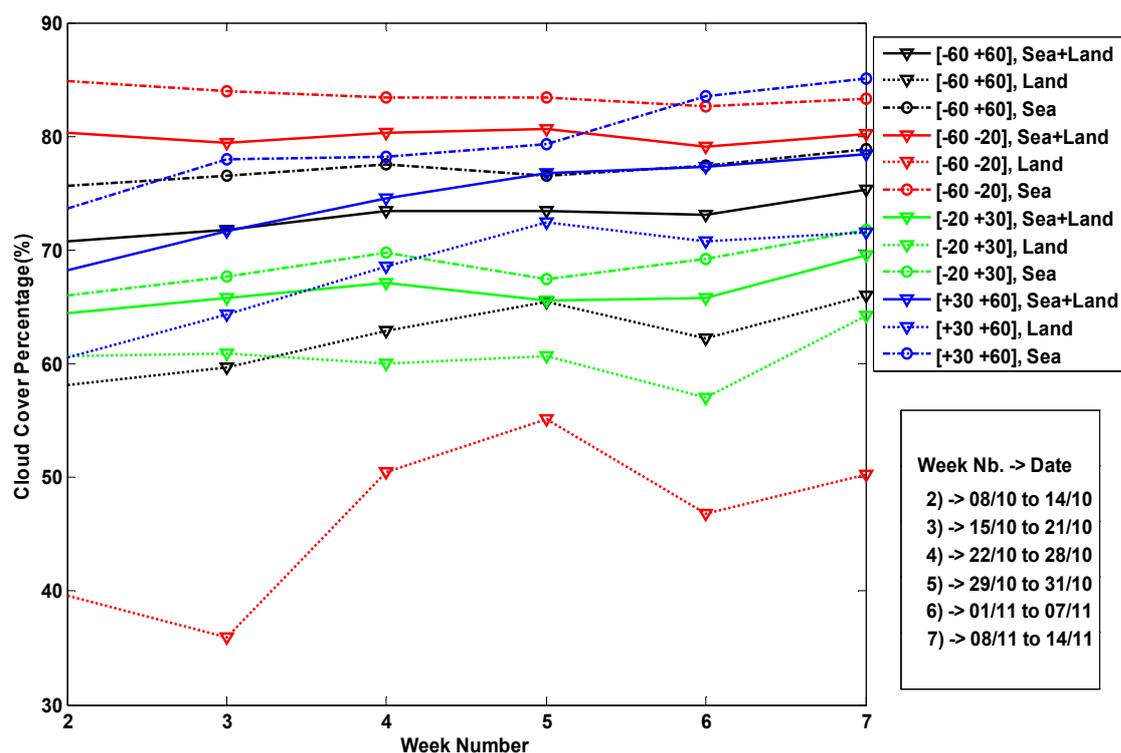
## 2.3.5 Résultats globaux obtenus à partir des données GLAS opérationnelles

L'orbitographie sur la première semaine d'octobre 2003 du satellite ICESat (voir la figure **Figure 2.3.24**), montre l'ensemble des  $15 \times 7 = 105$  révolutions orbitales correspondantes. Cette figure souligne bien l'homogénéité spatiale des enregistrements sur une telle période.



**Figure 2.3.24** : Orbites de la mission GLAS pour la période du 1 au 7 octobre 2003

Le couvert nuageux (voir la **Figure 2.3.25**), obtenu à partir des données GLAS sur les 6 semaines comprises entre le 6/10/2003 et le 14/11/2003 montrent une variation de 1 à 15 % suivant le type de surface continentale ou océanique considérée.



**Figure 2.3.25 :** Couverture nuageuse enregistrée par GLAS, sur chacune des semaines de la période comprise entre le 8/10/2003 et le 14/11/2003. Les résultats sont donnés sous forme globale et suivant les bandes de latitude, et suivant le type de surfaces (surfaces océaniques ou continentales) considérées.

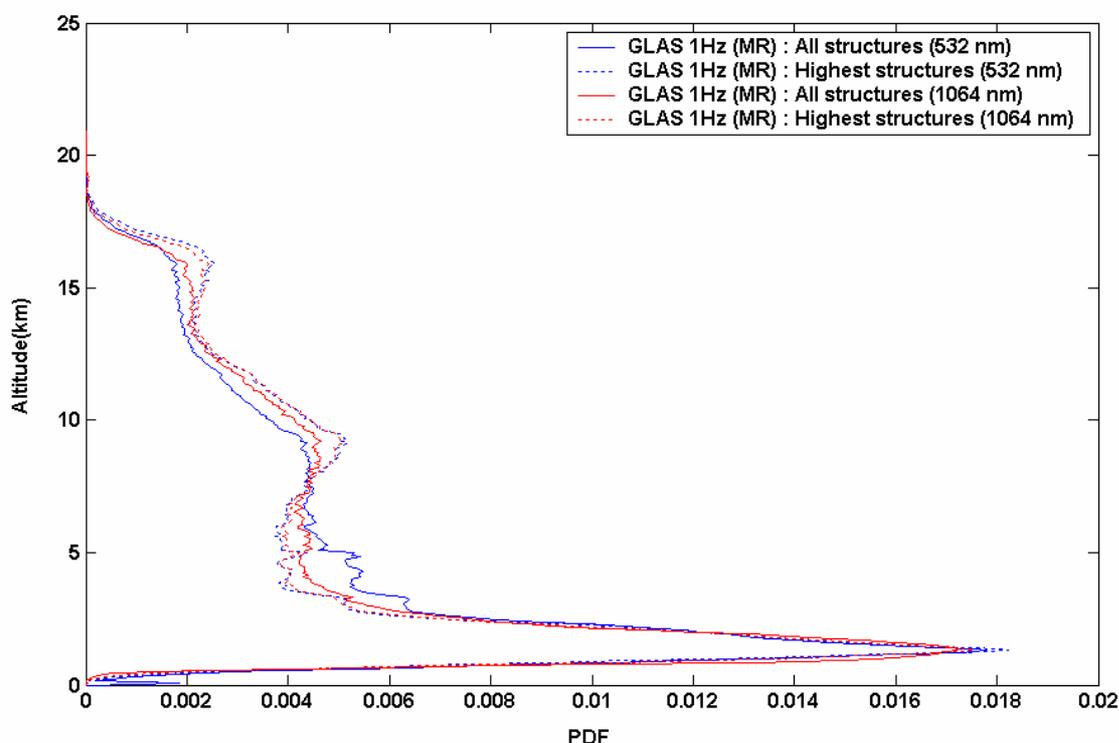
Nous pouvons à partir des données de la mission GLAS effectuer l'analyse statistique de l'altitude des sommets de structures nuageuses (voir la **Figure 2.3.26** et la **Figure 2.3.27**), à 532 et 1064 nm. On remarque tout d'abord de faibles différences entre ces PDF induites par la longueur d'onde.

On peut ensuite remarquer, grâce à l'étude de la **Figure 2.3.26**, une faible différence entre les PDF calculées en ne considérant pour chacun des tirs lidar que l'altitude de la structure nuageuse la plus élevée, ou en prenant en compte l'ensemble des sommets de structures détectées.

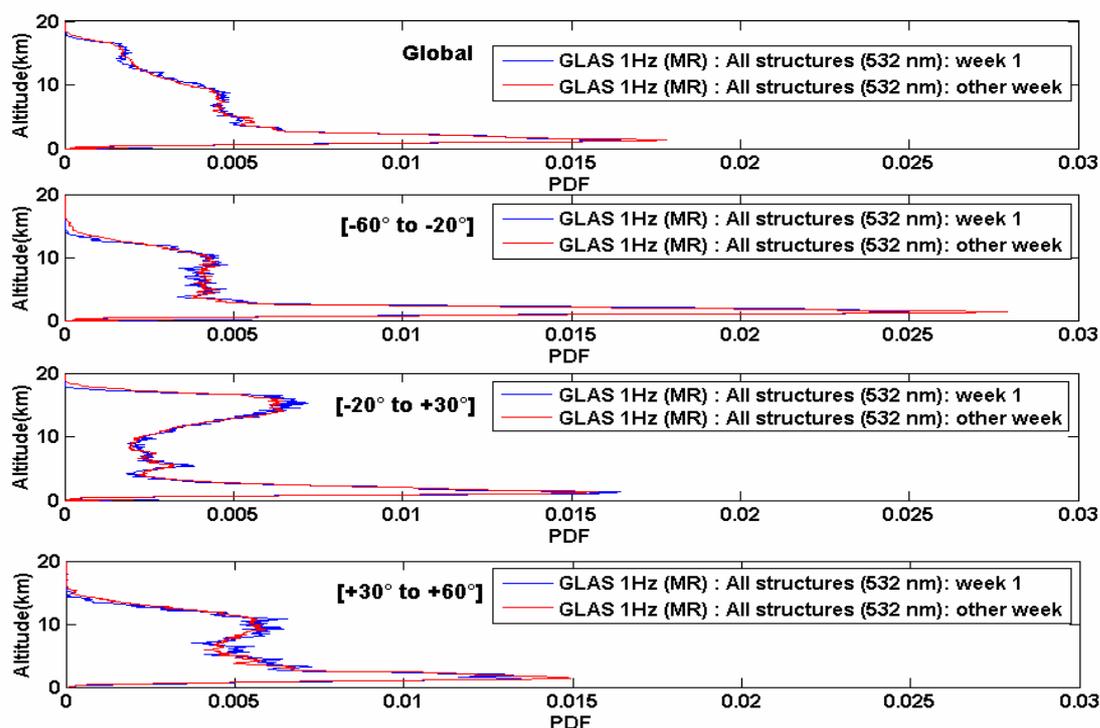
La raison en est que le rapport signal à bruit qui caractérise le signal de GLAS est faible. Ceci s'explique du fait que la distance entre le satellite et la terre est assez grande (590 km), pour une énergie d'émission du laser assez faible (36 mJ). Par conséquent, le signal lidar tombera assez rapidement dans le bruit lors de sa traversée des couches nuageuses optiquement denses. Après seuillage de ce signal par notre algorithme, nous allons donc par conséquent aboutir à des structures ayant une faible épaisseur géométrique. Avec ce même algorithme de seuillage, le nombre de tirs

montrant une structure multicouche sera aussi moins important que dans le cas de données lidars caractérisées par un SNR plus important.

Cette absorption est particulièrement visible sur la **Figure 2.3.27**. Une grande partie des structures nuageuses observées entre 3 et 10 km sur les bandes de latitudes [60°S ; 20°S] et [30°N ; 60°N], est occultée sur la bande de latitude [20°S ; 30°N]. En effet, cette zone de latitude montre la présence de nombreux nuages de type Cumulonimbus (Cb) reliés à la présence de l'ITCZ (Inter Tropical Convergence Zone), et caractérisés par une très forte extension verticale (jusqu'à environ 17 km) et une très grande épaisseur optique. La probabilité de détecter une structure nuageuse sous un tel type de structure est donc très faible.



**Figure 2.3.26 :** PDF de l'altitude du sommet des structures nuageuses sur l'ensemble de la mission GLAS, à 532 et 1064 nm, et à moyenne résolution (1Hz). Les PDF sont données soit en prenant en compte pour chaque tir l'ensemble des structures nuageuses (All), soit en ne conservant pour chaque tir que la structure nuageuse la plus haute (Highest structures). (pas d'intégration des PDF de 0.5 km).



**Figure 2.3.27 :** PDF de l'altitude des sommets de structures nuageuses, sur la première semaine et sur l'ensemble des semaines de la mission GLAS, à 532 nm et à moyenne résolution (1Hz). Les PDF sont calculées en ne conservant pour chaque tir que la structure nuageuse la plus haute (Highest structures). Cette étude a été effectuée sur les trois bandes de latitudes et en global (pas d'intégration des PDF de 0.5 km)

Une climatologie nuageuse en mode « tir à tir » est obtenue à partir des données GLAS, sur l'intervalle de latitude  $[60^{\circ}\text{S} ; 60^{\circ}\text{N}]$ . Pour l'ensemble des tirs lidar ayant, comme structure la plus élevée, une structure de haute altitude, nous donnons alors la répartition entre les classes de tirs (H), (H+M), (H+L) et (H+M+L) (voir la Table 2.3). Nous effectuons ensuite le même travail dans le cas des tirs lidar présentant comme structure la plus élevée une structure de moyenne altitude, et donnons comme précédemment la répartition entre les classes de tirs (M) et (M+L).

On note dans ces statistiques, une assez grande proportion de structures monocouches comparativement à l'ensemble des structure multicouches. En effet, les structures nuageuses monocouches classées comme moyennes (M), sont trois fois plus présentes (73%) que les structures multicouches (M+L). De même, considérant le cas des tirs comportant une structure de haute altitude, le pourcentage de cas monocouche (H) reste bien supérieur (57.3%) à l'ensemble des cas multicouches (H+M, H+L, L+M+H).

Highest		All latitude range	[-60°;-20°]	[-20°; +30°]	[+30°; +60°]
High	H	57.3 (57.6; 56.5) %	58.0 (57.8; 59.5) %	58.1 (58.2; 57.9) %	54.8 (55.5; 54.0) %
	H+M	18.6 (18.4; 19.1) %	18.6 (18.7; 17.8) %	18.4 (18.2; 18.7) %	19.2 (18.5; 20.0) %
	H+L	19.8 (19.7; 20.1) %	18.7 (18.7; 18.3) %	19.6 (19.8; 19.2) %	21.8 (21.9; 21.6) %
	L+M+H	4.3 (4.3; 4.3) %	4.7 (4.8; 4.4) %	3.9 (3.8; 4.2) %	4.2 (4.1; 4.4) %
		$\Sigma=100\%$	$\Sigma=100\%$	$\Sigma=100\%$	$\Sigma=100\%$
Middle	M	73.4 (73.4; 73.4) %	71.8 (71.8; 72.0) %	74.7 (75.2; 73.5) %	73.4 (73.2; 73.7) %
	M+L	26.6 (26.6; 26.6) %	28.2 (28.2; 28.0) %	25.3 (24.8; 26.5) %	26.6 (26.8; 26.3) %
		$\Sigma=100\%$	$\Sigma=100\%$	$\Sigma=100\%$	$\Sigma=100\%$

**Table 2.3** : Répartitions internes des classes nuageuses de GLAS pour lesquelles la structure la plus haute en altitude est de classe haute (H) ou de classe moyenne (M). Ces statistiques sont calculées dans l'intervalle de latitude [60°S ; 60°N]. Dans chacune des cases du tableau, le premier chiffre correspond à la couverture globale. Le second chiffre correspond à la couverture au dessus de la mer, et le troisième au dessus de la terre.

Ces observations nous permettent de vérifier la remarque effectuée plus haut dans cette section, sur le fait que le signal de GLAS, du fait de son faible SNR, tombe assez rapidement dans le bruit à la traversée des structures nuageuses denses. Cela nous permet aussi d'obtenir une information sur la proportion de nuages semi transparents dans chacune des classes. Les cas de tirs multicouches sont en effet un indice sur le fait que la structure d'altitude la plus élevée est de nature semi-transparente. La proportion de ces derniers paraît bien plus importante dans le cas des nuages de hautes altitudes (42.7% en global), que dans le cas des nuages de plus basses altitudes (26.6%). Il faut cependant tenir compte, pour caractériser la semi-transparence de chacune de ces classes d'altitude, de la proportion des tirs classés comme monocouches ((H), (M) et (L)), et dont la structure nuageuse est semi-transparente.

Cependant, cette information n'est pas disponible au sein de la base de données GLAS, telle que fournie dans le produit de niveau 2. Une première façon de se sortir de ce problème serait d'utiliser un niveau de traitement supérieur. Une autre manière de se sortir de ce problème serait

d'utiliser, après traitement des données GLAS par notre algorithme de seuillage, l'information liée à la détection, ou la non détection du sol. Nous pourrions ainsi conclure sur la nature semi-transparente ou dense de la structure. Nous utiliserons plus loin cette opportunité dans le cadre du traitement des données LITE.

Considérant les variations latitudinales, nous remarquons assez peu de différences dans la répartition interne des classes.

## 2.4 Application aux observations du lidar

### LITE

L'algorithme permettant l'extraction des données lidar satellite, décrit précédemment, a été construit à partir de l'étude des données de la mission GLAS. Afin de valider notre algorithme, et montrer sa robustesse en terme de performance, nous avons jugé nécessaire d'étendre son utilisation aux données lidar issues d'autres instruments de mesures satellite.

Hormis les données de CALIPSO non encore accessibles au moment de l'écriture de cette thèse, les données de la mission LITE sont les seules données disponibles afin de remplir cet objectif.

Nous allons donc, dans une première partie, adapter les paramètres de l'algorithme précédemment construits, afin que ces derniers puissent convenir aux caractéristiques des données de la mission LITE, et ceci pour les deux longueurs d'ondes 532 nm et 1064 nm.

Les données de la mission LITE diffèrent en effet grandement dans leur caractéristiques de celles de GLAS. Des différences importantes existent entre les capacités de l'instrument LITE (240 km d'altitude, une puissance émise à 532 nm (1064 nm) comprise entre 380 mJ et 530 mJ (400 et 471mJ), une ouverture de 3.5mrad de nuit/1.1mrad de jour), et celles de l'instrument GLAS (591 km d'altitude, une puissance émise à 532 nm(1064 nm) d'environ 36 mJ (74 mJ) , une ouverture à 532nm (1064 nm) de 0.19mrad (0.5 mrad)).

L'ensemble de ces différences intervient de façon conséquente sur l'acquisition d'une même scène nuageuse.

Le SNR présent sur les données LITE est bien supérieur à celui des données GLAS (d'un facteur 2 à 3). Par conséquent la détermination de la base des structures nuageuses, ainsi que la

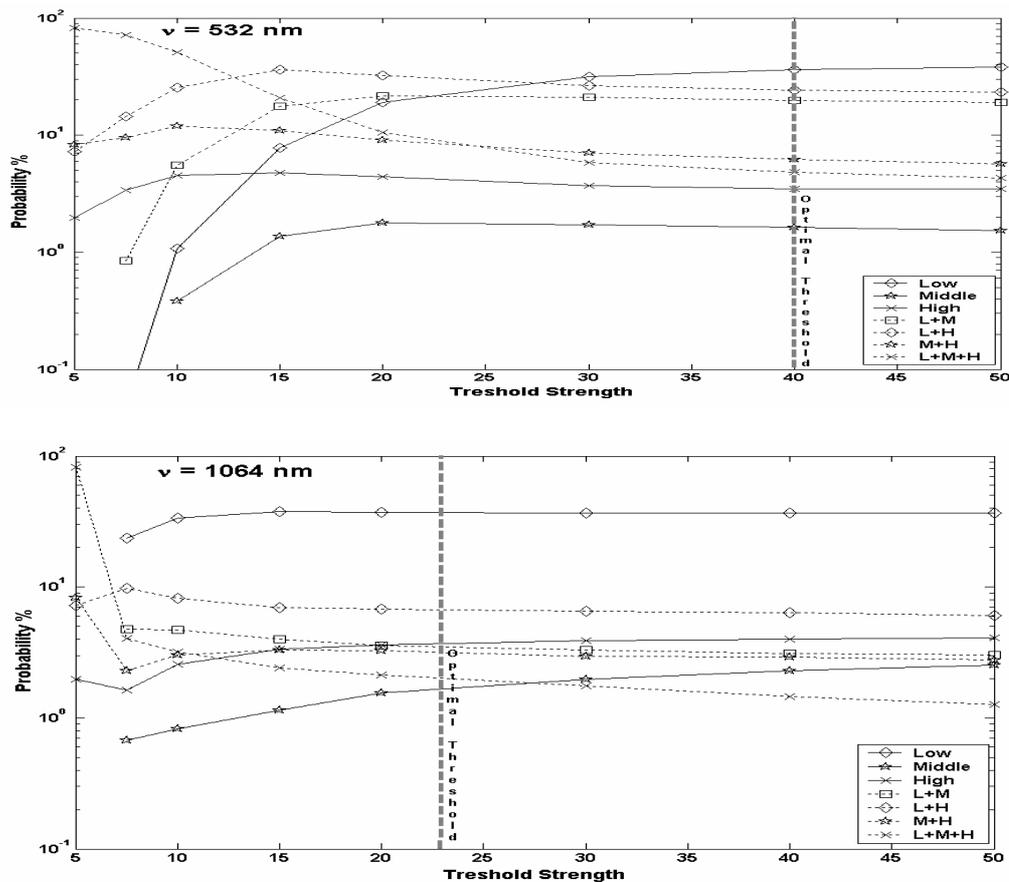
détermination des classes multicouches, seront par conséquent plus fortement biaisées dans le cas des données GLAS que dans le cas des données LITE.

On note aussi comme grande différence, une résolution verticale des données de la mission LITE (15 mètres), bien supérieure à celle fournie par la mission GLAS (76.8 mètres).

Ce travail d'adaptation de l'algorithme effectué, nous présenterons dans une seconde partie les résultats obtenus à partir du traitement des données LITE. Nous finirons ce chapitre sur l'intercomparaison des résultats obtenus grâce au traitement par notre algorithme de la base de données LITE, avec ceux obtenus à partir de la base de données GLAS traitées grâce à l'algorithme opérationnel.

## 2.4.1 Adaptation de la méthode

Nous allons tout d'abord, comme il a été fait précédemment dans le cadre du traitement des données GLAS, estimer pour les données LITE la valeur du facteur de force de seuillage  $F$  la plus appropriée afin de minimiser la probabilité de fausse détection. Nous estimons en fonction de la valeur de  $F$ , la répartition des classes nuageuses monocouches et multicouches (cf. la **Figure 2.4.1**).



**Figure 2.4.1 :** Fraction nuageuse en fonction de la force de seuil appliquée  $F$  à 532 nm (haut) et 1064 nm (bas).

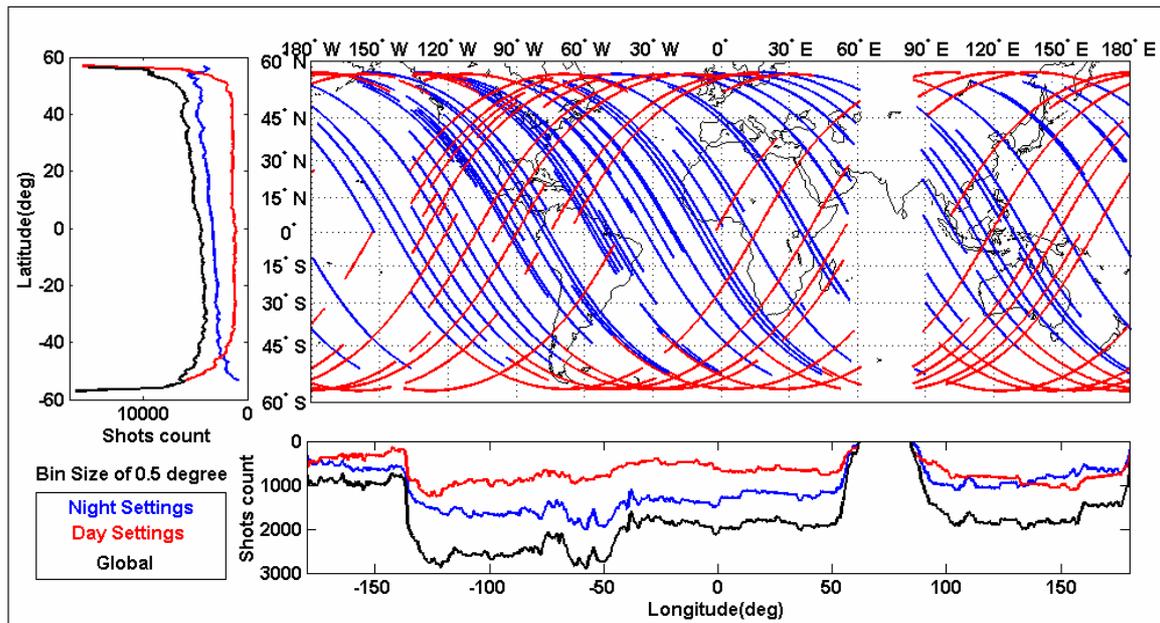
Pour les données prises à la longueur d'onde  $\lambda=532 \text{ nm}$ , une stabilisation des différentes couvertures nuageuses est obtenue pour des valeurs de  $F$  supérieures à 30. Pour de plus faibles valeurs de  $F$ , les occurrences de structures nuageuses monocouches de basse altitude décroissent rapidement, principalement au bénéfice des structures multicouches, indice d'un nombre croissant de structures bruitées. Une valeur de  $F$  équivalente à 40 semble être un bon compromis et a été adoptée par conséquent afin d'effectuer l'analyse des fractions nuageuses à 532 nm. La même étude de sensibilité est effectuée à 1064 nm, nous amenant à sélectionner une valeur plus faible pour  $F$ , et est égale à 23.

Pour ces valeurs de  $F$ , la sensibilité de la méthode en terme de probabilité d'erreurs sur la distribution des classes en fonction de la valeur choisie de  $F$  est d'environ 1%.

## 2.4.2 Résultats à partir des données LITE

### 2.4.2.1 Identification des structures verticales

Nous avons appliqué à l'ensemble des données de la mission LITE notre algorithme de seuillage, avec les valeurs de  $F$  telles que définies dans la section précédente.



*Figure 2.4.2 : Trajectographie de la mission LITE utilisée dans le cadre de l'analyse statistique. La densité en nombre de tirs suivant la latitude et la longitude est aussi donnée.*

Après le rejet de toutes les données considérées comme étant de qualité médiocre, il nous reste un ensemble d'environ  $7.6 \cdot 10^5$  profils lidar de nuit, qui ont été analysés avec l'algorithme précédemment décrit. La trajectographie totale correspondante représente 57 orbites, presque uniformément distribuées entre les latitudes  $40^\circ\text{S}$  et  $55^\circ\text{N}$ , bien qu'une grande part de l'océan pacifique et indien soit faiblement représenté (voir la **Figure 2.4.2**).

Les distributions retrouvées de tous les sommets de couches identifiés en fonction de l'altitude et de la latitude à 532 nm sont données dans la **Figure 2.4.3** et la **Figure 2.4.4**.

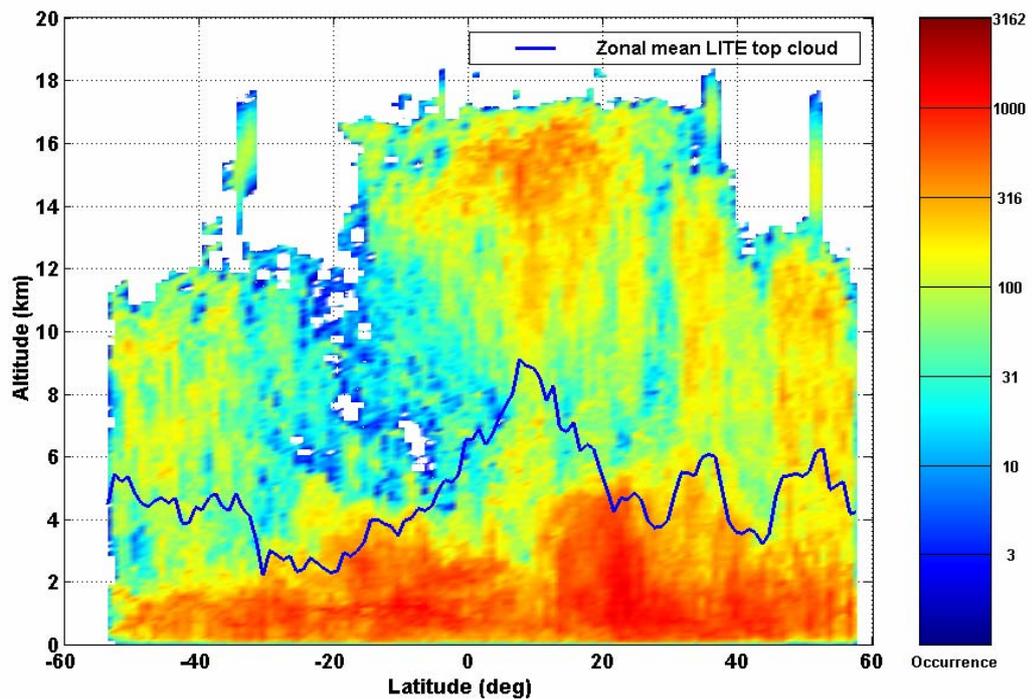


Figure 2.4.3 : Occurrence des sommets de structures détectées par notre algorithme de seuillage à la longueur d'onde  $\lambda=532$  nm.

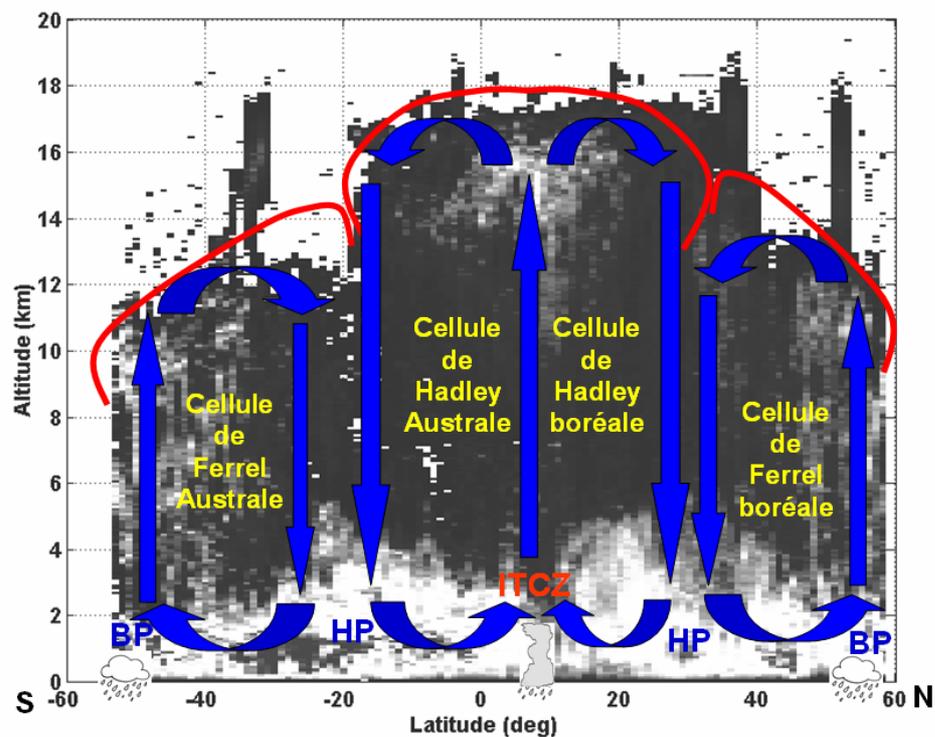


Figure 2.4.4 : Correspondances entre la circulation méridienne à grande échelle, et la distribution latitudinale des structures nuageuses/aérosols, telle que retrouvée grâce à notre algorithme de détection, appliqué aux données de la mission LITE.

On note tout d'abord sur cette figure que les altitudes maximales atteintes par les sommets de structures sont fonction de la latitude. Le comportement de la distribution de ces sommets de structure semble être différent, et peut être distingué suivant trois zones caractéristiques. Les intervalles [50°S-20°S] et [30°N-60°N], de moyenne latitude, montrent tous deux une altitude maximale d'environ 12 km, alors que la zone intertropicale [15°S-30°N], correspondant à l'ITCZ, montre pour sa part une altitude maximale atteinte d'environ 17 km, avec un maximum d'occurrence contenue entre 14 et 16 km. On retrouve ainsi une correspondance assez forte avec le schéma classique de la circulation méridienne atmosphérique à grande échelle. Ce dernier comporte six cellules de convection : deux cellules équatoriales, ayant un sens de circulation direct, dites cellules de Hadley, deux cellules à circulation inverse des précédentes dites cellules de Ferrel, et deux cellules polaires (à nouveau en circulation directe). Les cellules de Hadley et de Ferrel sont toutes deux mises en relief dans la **Figure 2.4.4**.

Du fait de la valeur de l'inclinaison de l'orbite de la navette Discovery lors de la mission LITE ( $i=57^\circ$ ), l'intervalle de latitude correspondant à l'orbitographie est compris entre 57°S et 57°N, et ne permet donc pas d'atteindre les cellules polaires.

L'ensemble de ces zones semble montrer une forte occurrence à 532 nm des sommets de structure pour les altitudes comprises entre 0 et 5 km. Ces occurrences semblent particulièrement importantes dans les zones de latitudes [20°S-0°N] et [15°N-35°N], et semblent être la signature de la présence d'aérosols.

Nous allons maintenant pouvoir nous intéresser aux statistiques concernant la couverture nuageuse, et la distribution des classes de tirs (voir la **Figure 2.4.5**). Cette étude est effectuée de manière séparée suivant les longueurs d'ondes  $\lambda=532$  nm et  $\lambda=1064$  nm. Nous allons ainsi montrer la dépendance de ces statistiques en fonction de la longueur d'onde  $\lambda$  choisie.

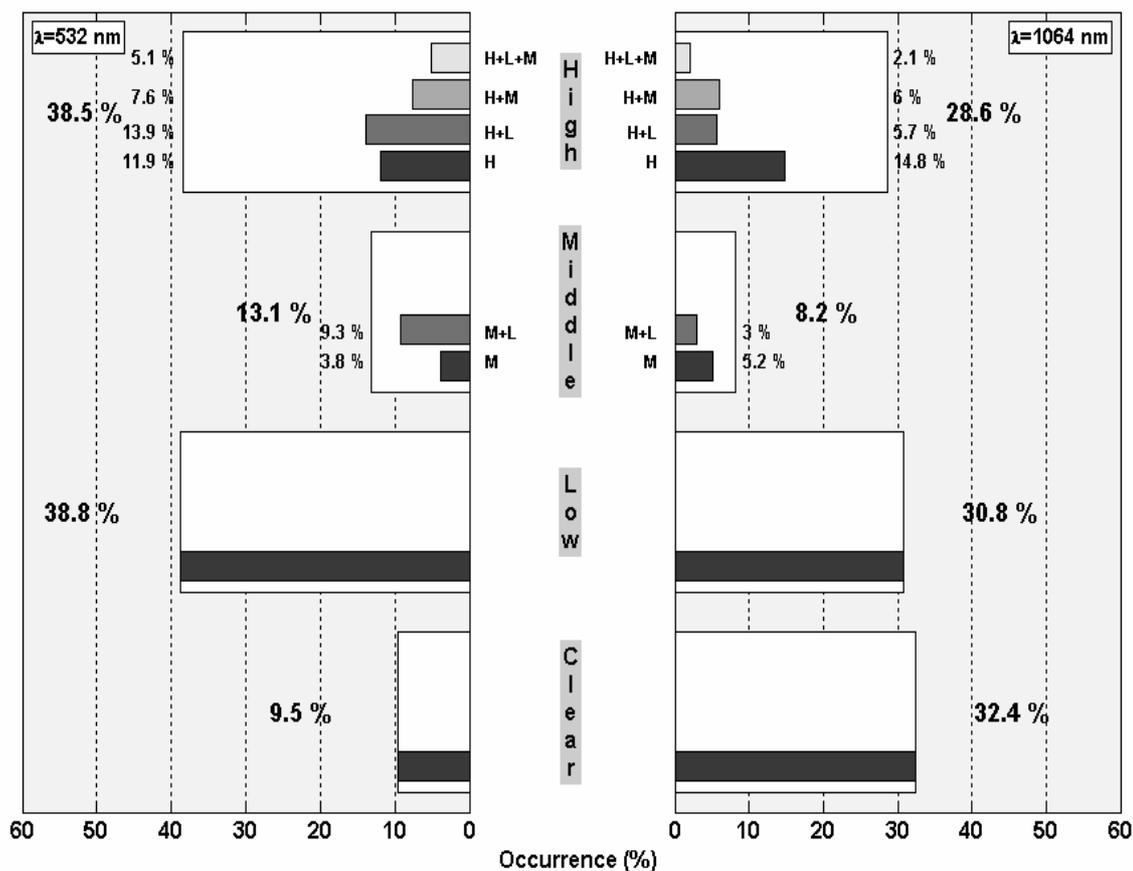


Figure 2.4.5 : Statistique de la couverture nuageuse et des distributions des structures obtenues à partir de l'analyse séparée de la base de données LITE à 532 et 1064 nm.

En utilisant la classification nuageuse ISCCP (B, M et H) pour les couches identifiées comme étant les plus hautes sur chacun des tirs lidar de LITE, nous obtenons à 532 nm (1064 nm) une occurrence de nuages hauts de 38.5% (28.6%), de nuages moyens de 13.1% (8.2%) et de nuages bas de 38.8% (30.8%). Les pourcentages pour chacune de ces classes nuageuses restent donc très comparables. Cependant, pour chacune d'elles, l'occurrence des couches multiples est significativement supérieure à 532 nm. On remarque aussi la différence entre ces deux longueurs d'ondes, quant au pourcentage de ciel clair, de 9.5% (32.4%) à 532 nm (1064 nm). Ces dernières observations vont dans le même sens que la précédente hypothèse, concernant une contamination, dans les statistiques nuageuses effectuées à 532 nm, par les couches aérosols.

## 2.4.2.2 Séparation des nuages et des aérosols

Nous avons mis en valeur dans la section précédente un comportement différent de notre algorithme de traitement suivant la longueur d'onde à laquelle on se trouvait (532 nm ou 1064 nm). Nous avons aussi mis en avant une contamination des résultats statistiques obtenus à 532 nm par la présence d'aérosols. La voie à 532 nm présente à la fois l'occurrence de couches nuageuses et de couches aérosols, alors que la voie à 1064 nm est exempte de structures aérosols.

Dans le cas particulier des poussières désertiques (appelées « Dust » en anglais), le coefficient d'Angstrom est connu pour être presque nul. Ces différences dans notre détection des structures aérosols de type désertique ne peuvent donc pas être expliquées par une dépendance spectrale du signal de rétrodiffusion  $\beta$ .

Or après vérification sur les données montrant des aérosols désertiques, nous avons bien détection par notre algorithme de ces derniers à 532 nm, alors qu'ils restent indétectables à 1064 nm.

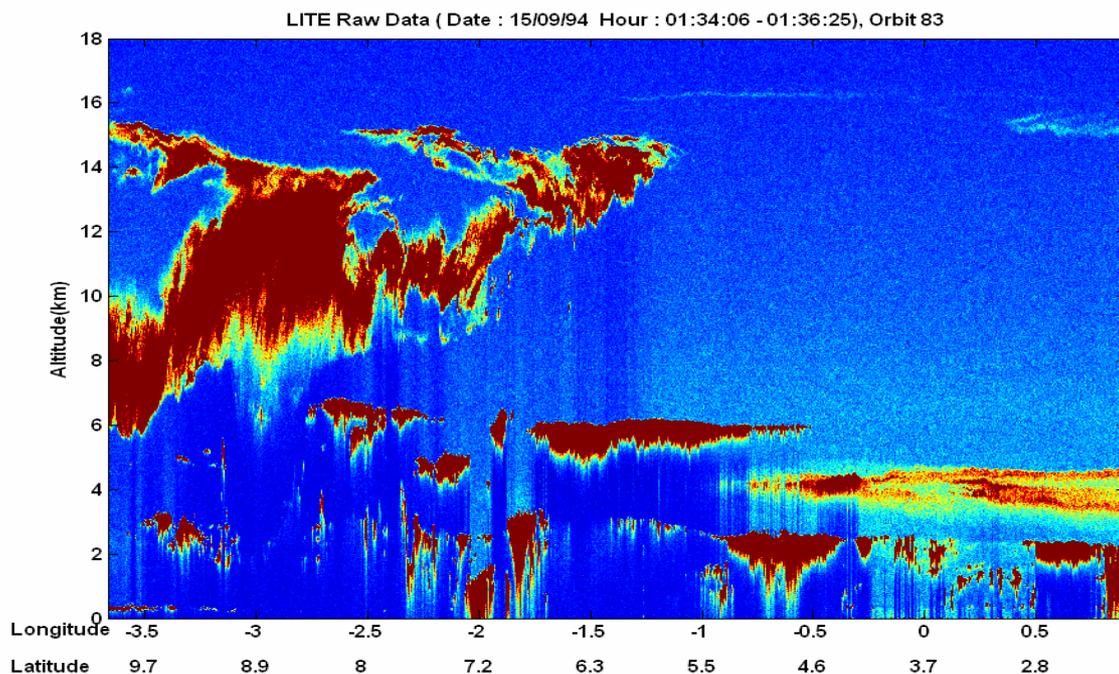
Après avoir étudié le signal lidar en présence d'aérosol, nous trouvons un rapport signal à bruit bien plus faible à 1064 nm, que dans le cas des données à 532 nm, ce qui explique les différences de comportement obtenues en présence d'aérosols.

Nous décidons alors d'utiliser la combinaison des deux canaux 532/1064nm afin d'effectuer la discrimination des structures nuageuses et aérosols. Pour les structures dont l'altitude du sommet est inférieure à 8.5 km, seules les couches détectées aux deux longueurs d'ondes sont alors identifiées comme étant des nuages. Au dessus de cette limite à 8.5 km, toutes les structures observées à 532 nm sont prises en compte, préférentiellement à celles observées à 1064 nm où l'épaisseur géométrique des couches nuageuses est sensiblement inférieure, indiquant une plus faible sensibilité.

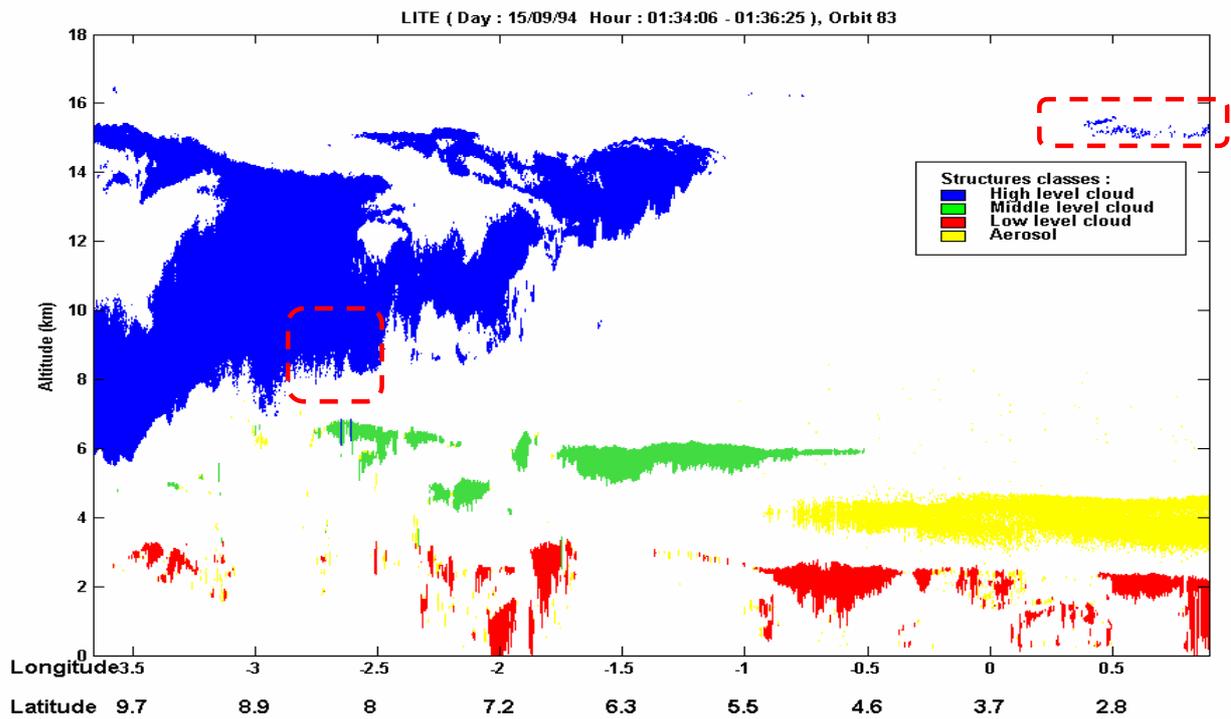
Il serait néanmoins intéressant, pour des travaux futurs, de montrer quelle est la différence entre cette discrimination nuage/aérosol basée sur la comparaison des structures obtenues à partir des deux canaux 532 nm/1064 nm, et la méthode utilisée par la science team de GLAS, qui est basée sur le seuillage du produit  $\beta \cdot d\beta$ .

La **Figure 2.4.6** (signal de rétrodiffusion brut) et la **Figure 2.4.7** (résultat de la classification nuageuse) présentent le résultat d'une telle classification effectuée à partir de la discrimination

nuage/aérosol, basée sur comparaison des structures détectées à 532 nm et 1064 nm. Le profil lidar présenté correspond à une section de l'orbite 83 de LITE, enregistrée le 15/09/1994, entre 01:34:06 (GMT) et 01:36:25 (GMT). Elle montre l'occurrence de structures nuageuses répartie sur les trois niveaux d'altitudes Bas (en rouge), Moyen (en vert) et Haut (en bleu). Les nuages hauts présents dans cette figure peuvent être distingués en deux types de nuages hauts bien caractéristiques. Tout d'abord un Cumulonimbus (Cb), dont la position est située entre  $4^{\circ}\text{W}$  et  $1^{\circ}\text{W}$ , et dont la structure massive et compacte, s'étend de 6 km jusqu'à environ 15.5 km d'altitude. Entre  $0.5^{\circ}\text{E}$  et  $1^{\circ}\text{E}$  et vers 16 km d'altitude, on observe la présence d'un cirrus, montrant une structure bien plus fractionnée, et dont l'épaisseur ne dépasse pas ici 500 mètres. Entre  $1^{\circ}\text{W}$  et  $1^{\circ}\text{E}$ , on observe entre 3.5 et 4.5 km d'altitude, la présence d'une structure, correctement classée, grâce à notre méthode de discrimination, comme étant une couche d'aérosols.

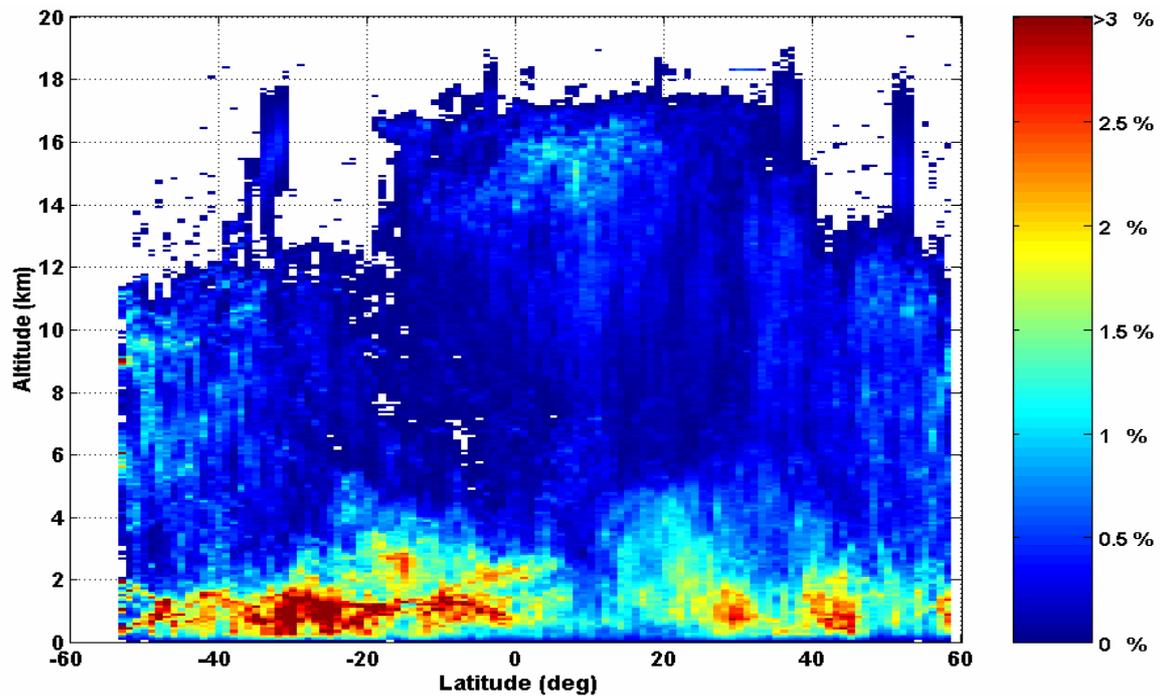


**Figure 2.4.6 :** Données lidar brutes correspondant à la section de l'orbite 83 de LITE, enregistrées le 15/09/1994, entre 01 :34:06 (GMT) et 01:36:25 (GMT).

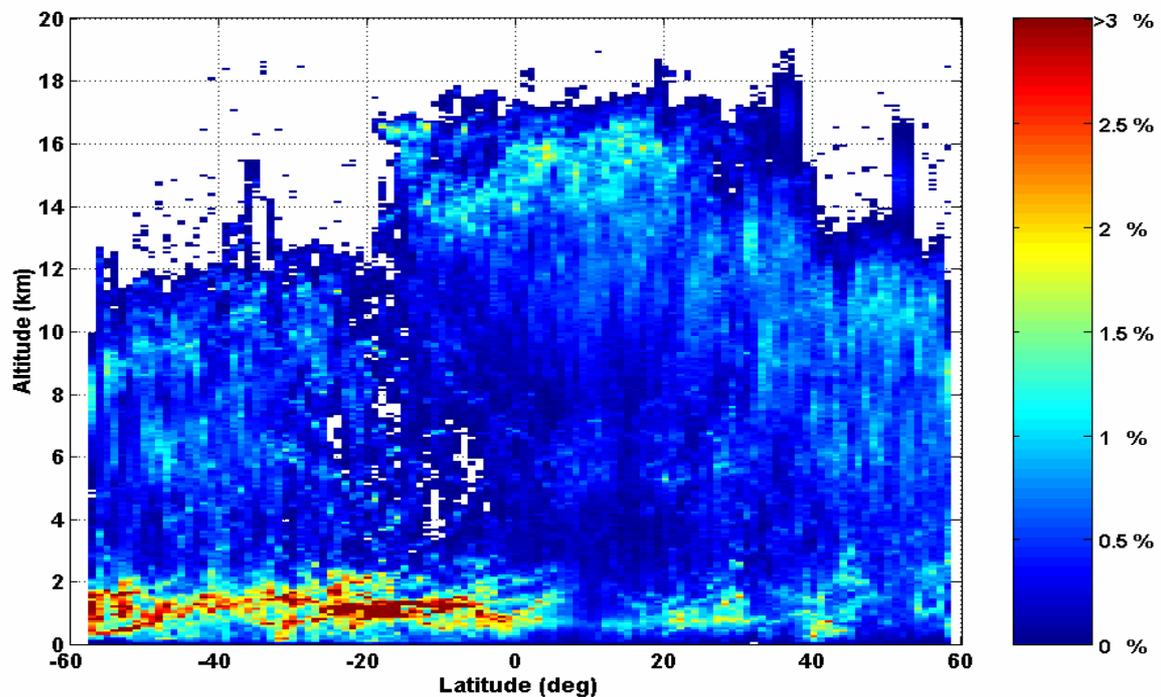


**Figure 2.4.7 :** Classification et identification des structures des aérosols et des nuages.

Cette dernière étape de l'algorithme d'extraction que constitue la discrimination entre les structures nuageuses et aérosols complétées, nous pouvons commencer la construction des statistiques des couches nuageuses de la mission LITE. La **Figure 2.4.8** et la **Figure 2.4.9** présentent la densité en latitude et en altitude de l'ensemble des sommets de structures détectées, avant et après avoir la discrimination nuage/aérosol.



**Figure 2.4.8 :** Densité des sommets de structures nuageuses et aérosols à 532 nm. La discrimination n'a pas encore été appliquée. La densité de structures est ici représentée proportionnellement au maximum de structures détectées avec la discrétisation utilisée, soit 3162 structures.



**Figure 2.4.9 :** Densité des sommets de structures nuageuses, après discrimination nuages/aérosols. L'échelle de couleur est la même que celle utilisée dans la **Figure 2.4.8**.

La comparaison entre les distributions latitudinales des structures nuageuses, avant et après discrimination met en relief l'influence significative de ces structures aérosols pour une étude statistique de la basse et moyenne troposphère.

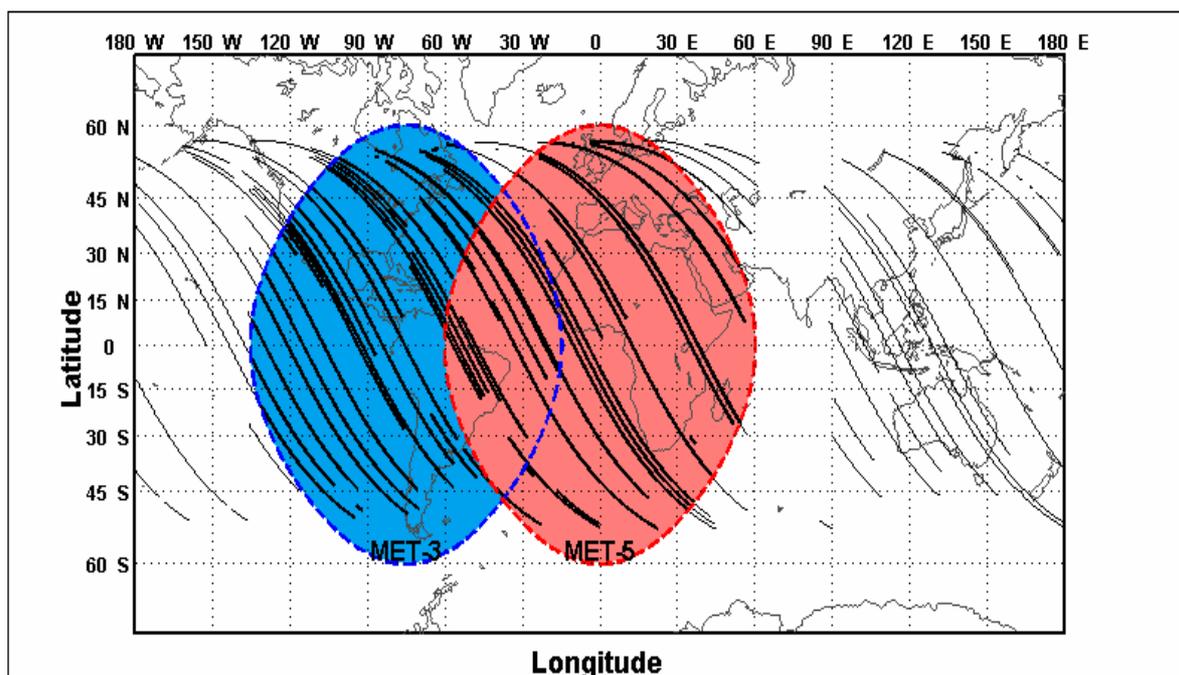
Les modifications les plus importantes, correspondant à la présence d'aérosols sont :

- **Entre 10°N et 35°N** : Une occurrence des aérosols majoritairement due à la présence de poussières désertiques dans la zone Saharienne, ainsi qu'une contribution très importante du panache est-ouest issue de cette zone. La péninsule Arabo-Persique, le Moyen et le proche Orient sont aussi des régions qui semblent montrer une occurrence importante de ces aérosols désertiques.
- **Entre 10°S et 40°S** : Une présence majoritairement due à l'occurrence de feux de biomasses (Biomass burning), générés au dessus de l'Amérique du sud, du bassin Amazonien et de l'Afrique du Sud.

L'étude que nous venons de commencer sur la distribution géographique de ces structures aérosols qui sont détectées grâce à cette méthode de discrimination, fera l'objet du chapitre 4 de cette thèse.

### **2.4.2.3 Analyse statistique en mode tir à tir sur les structures nuageuses**

Une analyse statistique de la distribution du sommet des structures nuageuses exclusivement, à partir des données nocturnes de la mission LITE peut maintenant être effectuée. Cette analyse est dans un premier temps présentée en mode « tir à tir » (voir la **Table 2.4** et la **Table 2.5**). Chacun des tirs lidar est classé suivant l'altitude de la structure nuageuse détectée la plus élevée. Comme dans le cas des statistiques GLAS, les classes d'altitudes sont identiques à celles utilisées dans le cadre de ISCCP, c'est-à-dire suivant les classes de nuages Bas (L), Moyen (M) et Haut (H), séparées par les niveaux de pression 680 et 440 hPa. On effectue aussi comme précédemment au sein de chacune de ces classes un classement interne prenant en compte l'information multicouche.



**Figure 2.4.10 :** Orbits de la mission LITE (orbites nocturnes) utilisées dans cette étude. La surface en bleue (rouge) correspond à la couverture des données ISCCP à la résolution DX données par le satellite Meteosat-3 (Meteosat-5).

Les résultats de cette analyse de la distribution nuageuse seront par la suite comparés avec ceux obtenus à partir d'autres détecteurs, et entre autres les détecteurs passifs de la base de données ISCCP (cf. Chapitre 3). Les données LITE seront dans un premier temps comparées aux données ISCCP-D1 et ISCCP-DX. Bien qu'offrant l'avantage de bénéficier d'une couverture globale, les données ISCCP-D1 sont caractérisées par une résolution spatiale de  $280 \times 280 \text{ km}^2$ , difficilement comparable avec celle de LITE (résolution de 740 m sur l'horizontale). Le choix a donc été fait d'effectuer en parallèle une comparaison avec les données ISCCP-DX, ayant une résolution spatiale plus appropriée à notre étude ( $30 \times 30 \text{ km}^2$ ). Cependant, nous n'avons eu accès à cette dernière résolution qu'aux données des satellites Meteosat-5 et Meteosat-7 de ISCCP (Les données des satellites NOAA-11 et NOAA-12 n'ayant pas été récupérées). La zone de couverture des données ISCCP-DX se trouve donc réduite par rapport à celle de ISCCP-D1 (cf. la **Figure 2.4.10**). L'analyse statistique des observations LITE a donc été dissociée (voir la **Table 2.4**) suivant les zones de couverture respective de ISCCP-D1 (chiffres de couleur noir) et de ISCCP-DX (chiffres de couleur rouge).

Highest		All latitude range	[-60°;-20°]	[-20°;+30°]	[+30°;+60°]
High	H	48.3 % 48.0 %	36.2 % 39.7 %	52.8 % 52.3%	46.9 % 45.4 %
	H+M	21.5 % 21.5 %	23.9 % 21.2 %	17.6% 18.3 %	26.7 % 27.5 %
	H+L	23.7 % 23.7 %	30.3 % 29.4 %	24.9 % 24.7 %	18.2 % 17.7 %
	H+M+L	6.5 % 6.8 %	9.6 % 9.7 %	4.7 % 4.7 %	8.2 % 9.4 %
Middle	M	62.2 % 63.0 %	48.0 % 47.5 %	69.7 % 73.1 %	63.7 % 64.5 %
	M+L	37.8 % 37.0 %	52.0 % 52.5 %	30.3 % 26.9 %	36.3 % 35.5 %

**Table 2.4 :** Répartitions nuageuses dans chaque classe d'altitude de LITE en mode « tir à tir » sous la couverture ISCCP-D1 (chiffres noirs) et ISCCP-DX (chiffres rouges), en mode tir à tir.

Highest	All latitude range	[-60°;-20°]	[-20°;+30°]	[+30°;+60°]
<b>High</b>	52.4% 49.8 %	37.6 % 41.0 %	57.4 % 51.3 %	56.1 % 55.3 %
<b>Middle</b>	9.0 % 9.2 %	10.2 % 9.9 %	6.6 % 6.7%	12.4 % 13.8 %
<b>Low</b>	38.6 % 41.0%	52.2 % 49.1 %	36.0 % 42.0 %	31.5 % 30.9
<b>Cover Ratio</b>	69.8 % 69.0 %	72.8% 78.5 %	72.5 % 70.2 %	64.0 % 59.7 %

**Table 2.5 :** Climatologie nuageuse LITE en mode « tir à tir ». Les chiffres en noir correspondent à la couverture des données ISCCP-D1, les chiffres en rouge correspondent à celle des données ISCCP-DX.

Les résultats obtenus en mode « tir à tir » (voir la **Table 2.4** et la **Table 2.5**) montrent l'importance significative de l'occurrence des tirs présentant une structure multicouche. Globalement, en mode « tir à tir », 51.7 % des tirs dont la classe la plus haute est de type **H** montre une structure multicouche. Cette proportion est de 37.8 % pour les classes de type **M**. Cette information essentielle sur la structuration multicouche, apportée à l'échelle globale par l'utilisation seule de la télédétection active, nécessite d'être prise en compte dans l'ensemble des modèles de climatologies nuageuses, car elle a un impact important sur le bilan radiatif.

En effet, à la différence d'un détecteur actif qui peut donner de façon directe la distribution des couches nuageuses dans la colonne atmosphérique, l'estimation des caractéristiques d'une structure nuageuse par un détecteur passif s'effectue de façon indirecte. Leur altitude est estimée à partir de la température de brillance. Sans autre information, les algorithmes de traitements utilisés font alors l'hypothèse de structures nuageuses monocouches, dont l'altitude correspond à l'altitude d'un nuage fictif radiativement équivalent à l'ensemble de la colonne atmosphérique.

Le lidar spatial est le seul instrument nous apportant cette information sur la distribution nuageuse. De multiples biais peuvent toutefois, il est vrai, être introduits par l'utilisation de la seule base de données LITE.

Un des biais existant réside dans le fait qu'à la différence d'un système de détection passif, offrant une couverture globale et très répétée temporellement, la mission spatiale LITE, dont la durée totale des enregistrements n'a pas excédé 53 heures, offre un maillage dont l'homogénéité peut être critiquée du point de vue de la représentativité géographique. Ce biais est aggravé par la

zone de longitude non couverte par LITE, correspondant au continent Indien (zone non enregistrée due à une coupure de l'accès au relais satellite TDRS (Tracking and Data Relay Satellite System) correspondant à cette zone pour la mission LITE).

	<b>Global</b> <b>[60°S ;60°N]</b>	<b>[60°S ;20°S]</b>	<b>[20°S ;30°N]</b>	<b>[30°S ;60°N]</b>
<b>LITE</b>	65 %	82 %	67 %	51 %
<b>GLAS</b> <b>(01 au 07/10)</b>	72 %	89 %	71 %	48 %
<b>MODIS</b>	74 %	91 %	74 %	51 %
<b>Réalité</b>	74 %	91 %	74 %	51 %

**Table 2.6 :** Pourcentage des observations effectuées au dessus de la mer sur le nombre total de données, suivant le type d'instrument (LITE, GLAS (première semaine) et MODIS) utilisé, en global et par intervalles de latitude. La dernière ligne du tableau donne la surface au sol réelle.

Cet aspect est illustré ici dans la **Table 2.6**, où le rapport entre le nombre de données enregistrées au dessus des surfaces océaniques sur le nombre total de données montre en ce qui concerne LITE d'assez fortes différences (9% en global) avec les pourcentages réels dus à la distribution géographique naturelle des surfaces continentales et des surfaces océaniques. Cette différence paraît d'autant plus grande que la zone à étudier est haute en latitude.

Le comportement nuageux au dessus des continents et au dessus des océans étant différent, ce biais aura de facto un impact dans la distribution globale nuageuse. Ce dernier reste néanmoins difficilement quantifiable.

Cette différence de représentativité est fortement gommée dans le cas de l'utilisation des données GLAS (2% en global), de part une durée de mesures considérablement accrue (1mois de mesure). Ce biais pourra par conséquent être négligé dans le cadre des futures missions spatiales, telles que celles de CALIPSO, qui sont prévues pour offrir une bien plus grande couverture temporelle et spatiale.

### 2.4.3 Intercomparaison des statistiques LITE et GLAS (données opérationnelles)

Malgré les différences, répertoriées dans la section précédente, dans l'acquisition des données LITE et GLAS, les résultats statistiques obtenus peuvent néanmoins être intercomparés. Cette étape essentielle de comparaisons entre les résultats obtenus par ces deux instruments de détection active, permet dans une certaine limite, la validation des distributions de structures nuageuses précédemment retrouvées.

Cette connaissance de la structure nuageuse définie, nous effectuerons par la suite la comparaison des climatologies nuageuses avec celles obtenues grâce aux détecteurs passifs (voir chapitre 3). Nous identifierons ainsi les caractéristiques propres à chaque instrument, et verrons comment un couplage actif/passif peut être mis en place afin d'améliorer notre connaissance de ces climatologies nuageuses. Cette étude trouve sa place dans le cadre des missions spatiales, qui offrent aujourd'hui, comme la mission CALIPSO/AquaTrain, et offrirons dans le futur de telles opportunités de couplages.

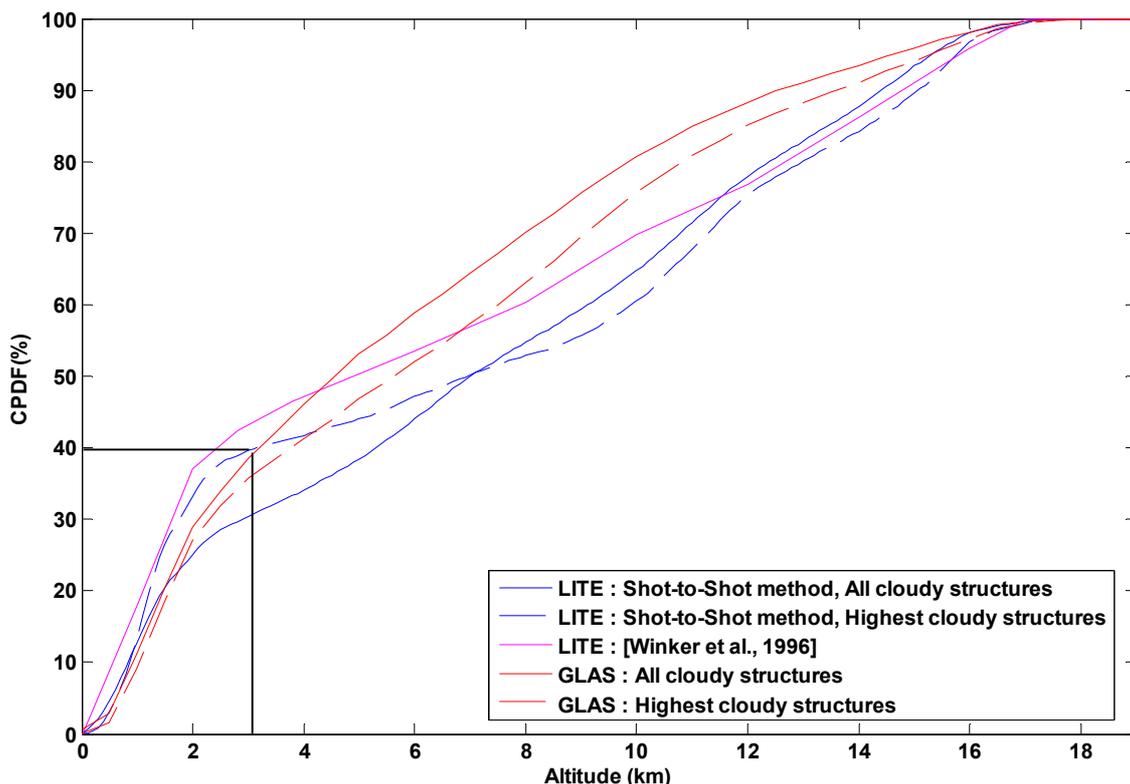
L'ensemble des statistiques de GLAS et LITE étant établies, nous cherchons donc à faire la comparaison de ces résultats. Afin de mener cette comparaison, nous cherchons donc dans un premier temps à nous affranchir des problèmes liés à la structuration multicouche. En conséquence, l'altitude du sommet de la structure nuageuse la plus haute sur chaque tir lidar nous semble être le paramètre le plus judicieux à utiliser afin de commencer cette comparaison.

Il est intéressant de comparer la forme des PDF cumulées (CPDF) obtenues à partir des deux jeux de données LITE et GLAS (voir la **Figure 2.4.11**). On remarque tout d'abord que [Winker et al., 1996] a fourni très rapidement après la mission LITE une CPDF du sommet des structures nuageuses les plus hautes sur chacun des tirs lidar.

Cette dernière est très proche de la CPDF que nous avons calculé dans les mêmes conditions grâce à notre algorithme d'extraction sur les données LITE. Nous fournissons aussi la CPDF correspondant à l'ensemble des structures présentes sur chacun des tirs lidar.

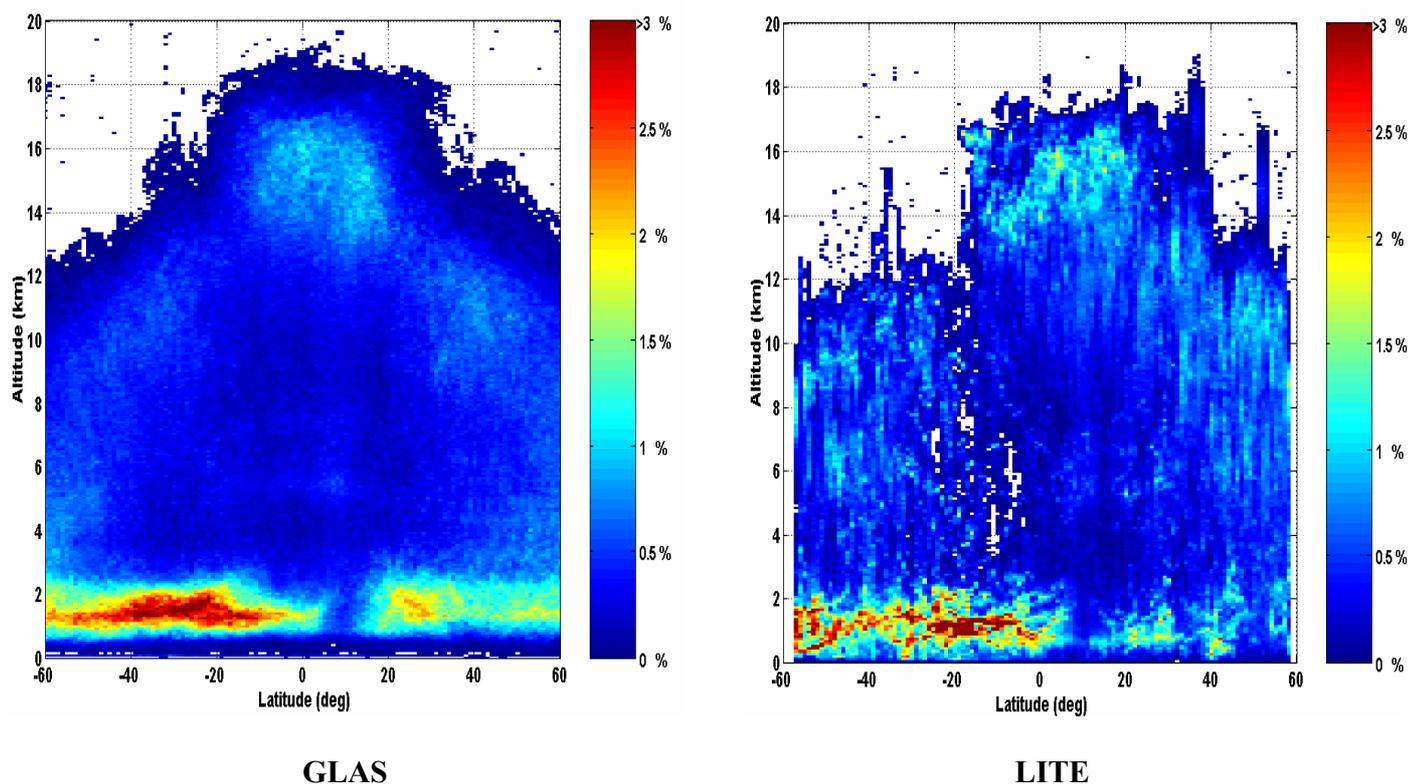
Comme le montre aussi cette figure, on note une très forte similitude, dans les CPDF issues des données LITE et GLAS. Cette quasi-similarité entre les résultats constitue une première preuve de la cohérence des statistiques calculées à partir de ces deux instruments.

La forme des CPDF de LITE et GLAS correspondant au sommets de structures les plus élevés sur chacun des tirs lidar (Highest cloudy structures), montre une cassure autour de 3 km d'altitude. On note qu'en ce point, la CPDF de LITE (GLAS) montre une probabilité cumulée de 40%(35%). Cette observation correspond avec les statistiques précédemment établies par Warren et al. (1985), qui estiment à 40% le taux de nuages bas non recouvert.



**Figure 2.4.11 :** CPDF de l'altitude des sommets de structures nuageuses, obtenues à partir des instruments actifs LITE (532 nm) et GLAS (532 nm à 1Hz). Les CPDF sont calculées pour toutes les structures nuageuses détectées et pour les structures détectées les plus hautes. Les résultats obtenus par [Winker et al.,1996] sur les données LITE sont aussi présentés.

Mais cette analyse reste une analyse globale. Il est donc nécessaire de vérifier dans un deuxième temps si les similitudes observées sont conservées suivant la latitude. On trace pour cela la distribution bidimensionnelle (latitude et altitude) de l'ensemble des structures nuageuses détectées (voir la **Figure 2.4.12**).



**Figure 2.4.12 :** Distributions latitudinales et en altitude des sommets de structures, retrouvées à partir des données des instruments GLAS (Première semaine) et LITE, en mode « tir à tir ». Les échelles de couleurs sont des échelles relatives, correspondant au nombre maximum de structures détectées dans chaque distribution.

Bien que la distribution de GLAS soit moins bruitée que celle de LITE, on observe une fois de plus la forte ressemblance entre les distributions retrouvées. Les structures nuageuses de faibles altitudes (entre 1 et 3 km) semblent être correctement retrouvées, ainsi que le positionnement en altitude et latitude de l'ITCZ. On retrouve le schéma de circulation atmosphérique méridienne à grande échelle précédemment dans la **Figure 2.4.4**.

Comme nous l'avons dit précédemment dans la section 2.5.5, le fait que le SNR de LITE soit supérieur à celui de GLAS, implique que le signal, à la traversée des couches nuageuses, chute plus rapidement dans le bruit. Par conséquent, le nombre de tirs classés multicouches sera aussi, par conséquent, plus faible dans le cas des données GLAS que dans le cas des données LITE. On observe ainsi sur la **Figure 2.4.12**, dans l'intervalle de latitude [20°S ; 30°N] et entre 3 et 10 km d'altitude, une baisse significative du nombre de structures nuageuses détectées par GLAS, comparativement à celles détectées par LITE.

Après avoir montré que les CPDF des structures nuageuses les plus élevées sur chacun des tirs lidar sont similaires entre les bases de données LITE et GLAS, et après avoir montré ici que les

distributions de l'altitude de ces sommets de structures élevées suivant la latitude restent aussi comparables, on va maintenant considérer dans la suite de ce travail l'ensemble des structures nuageuses présentes sur chacun des tirs lidar. Il nous faut donc montrer, pour clore cette comparaison entre les statistiques LITE et GLAS, que la structuration multicouche à l'intérieur de chacune des classes d'altitude reste identique. Afin de remplir cet objectif délicat, nous comparons les climatologies nuageuses globales de LITE (données dans la Table 2.4), avec celle de GLAS (données dans la Table 2.3). Pour plus de facilité, ces statistiques sont redonnées au lecteur dans la Table 2.7. Lorsque l'on compare les statistiques de LITE, et celles de GLAS, on remarque que toutes latitudes confondues, leurs valeurs restent dans le même ordre de grandeur (à  $\pm 10\%$ ). Ces écarts restent tout de même acceptables aux vues des différences d'acquisitions précédemment évoquées entre les deux bases de données.

A partir de la **Table 2.7**, nous calculons pour LITE et pour GLAS, le rapport du pourcentage de tirs monocouches sur le pourcentage de tirs multicouches, et ceci pour chacune des classes d'altitude haute et moyenne.

		LITE		GLAS	
		Occurrence des structures	Rapport monocouches/ multicouches	Occurrence des structures	Rapport monocouches/ multicouches
Classe Haute	H	48.3 %	0.93	57.3 %	1.34
	H+M	21.5 %		18.6 %	
	H+L	23.7 %		19.8 %	
	H+M+L	6.5 %		4.3 %	
Classe Moyenne	M	62.2 %	1.64	73.4 %	2.75
	M+L	37.8 %		26.6 %	

**Table 2.7 :** Rapport de l'occurrence des structures monocouches sur celle des structures multicouches, pour les missions LITE et GLAS.

La proportion des cas monocouches/multicouches à l'intérieur de chaque classe d'altitude semble néanmoins être conservée. On peut conclure de cette analyse que le rapport de nuages monocouches hauts sur les cas de nuages multicouche est compris entre 0.93 et 1.34, et que ce même pourcentage est compris entre 1.64 et 2.75 dans le cas des classes de nuages moyennes.

On observe que le nombre de tirs contenant en même temps les trois classes d'altitude haut, moyen et bas est compris entre 4.3 et 6.5%.

A la différence des données LITE, on observe cependant dans les données GLAS, peu de différences statistiques entre les structures nuageuses obtenues au dessus des surfaces maritimes et celles obtenues au dessus des surfaces continentales (voir la **Table 2.3**). Ces différences peuvent néanmoins être expliquées par le faible nombre de données de LITE comparé à celui de GLAS. Ces différences sont d'autant plus accrues lorsque l'on restreint la statistique de LITE à un type de surface (continentale ou maritime), ce qui réduit encore le nombre de données statistiquement disponibles. On souligne encore ici l'apport de la mission GLAS et les futures missions spatiales lidar telles que la mission CALIPSO, en terme de représentativité statistique spatiale et temporelle.

Cette intercomparaison tend à montrer la validité des climatologies nuageuses ainsi retrouvées, et par la même, la validité des algorithmes de traitements mis en place.

## **2.5 Etude de la structure de l'objet nuageux**

### **2.5.1 Positionnement du problème (qu'est-ce qu'un nuage ?)**

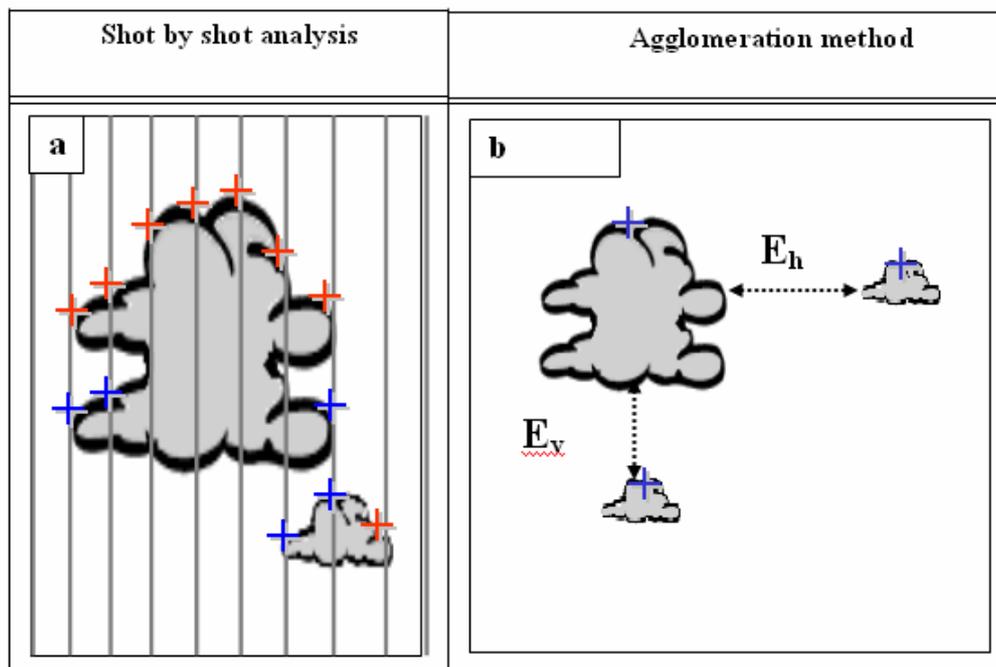
Les instruments de détections qui nous permettent généralement d'effectuer les mesures de la structure nuageuse, sont caractérisés par des résolutions verticales et horizontales, ainsi que par des méthodes de mesures, qui peuvent être très différents.

Pour ce qui est des résolutions, ces dernières induisent des modifications dans la perception que nous pouvons avoir de ce qui constitue une structure nuageuse. Il est bien évident, en effet, qu'un nuage sera vu différemment suivant que celui-ci est perçu par l'œil, par un radiomètre passif, ou par un instrument de télédétection actif. De plus, pour chacun de ces détecteurs, la détection d'une même scène nuageuse sera biaisée différemment suivant le paramètre considéré (l'altitude, l'épaisseur, la longueur, le nombre de structures...).

Ces différences de perceptions sont directement reliées à une perte d'information statistique, qui est due de façon formelle au non respect de la limite d'échantillonnage introduite par le théorème de Shannon [Shannon, 1948, 1949]. Pour rappel, ce dernier énonce que pour avoir un échantillonnage sans pertes d'informations, la fréquence de l'échantillonnage d'un signal doit au moins être égale au double de la fréquence de ce dernier. Par conséquent, le pas (équivalent à la taille des pixels) qui doit être utilisé pour effectuer l'échantillonnage doit être inférieur à la moitié de la taille du plus petit détail de l'image.

Un cas simpliste de forme de structures nuageuses servira la suite de notre explication (voir la Figure 2.5.1 (a) et (b)).

Les interprétations statistiques des matrices binaires de détections des nuages, obtenues en sortie de notre algorithme de seuillage, devront donc tenir compte des résolutions instrumentales utilisées. Les analyses directes des matrices binaires de détections ont été effectuées jusqu'à maintenant dans le cadre des données LITE et GLAS, à la résolution la plus importante. Notre façon de percevoir les structures nuageuses s'est donc limitée à une vision en mode « tir à tir », sans autres informations apportées sur la dimension horizontale. C'est le cas de l'ensemble des analyses de structures nuageuses faites à ce stade de notre étude.



**Figure 2.5.1 :** Schémas des différentes configurations de statistiques possibles. A gauche, une illustration de l'analyse en mode tir à tir. A droite, l'illustration correspondante à la reconnaissance de forme et à l'agglomération des structures.

Cependant, l'analyse d'un cas simple de nuage présentant deux «étages» d'altitude (voir la **Figure 2.5.1**), permet de démontrer que la détermination du paramètre de sommet de structure nuageuse sera biaisée sans autres informations sur la dimension horizontale. En effet, dans un tel mode de vision « tir à tir », l'ensemble des croix bleues (sommet des structures recouvertes par une autre structure nuageuse), ainsi que l'ensemble des croix rouges (sommets les plus élevés sur le tir lidar considéré) seront associées indifféremment à l'ensemble plus général des sommets de structures. Appliqué à l'exemple simple de la **Figure 2.5.1(a)**, nous arriverions à la conclusion erronée de la présence de deux structures nuageuses superposées pour un même nuage.

Une analyse plus « intelligente » de la structure sur sa dimension verticale et horizontale nous aurait en effet permis de nous rendre compte que ces deux étages ne constituaient en réalité qu'un seul et même nuage, et que celui-ci ne devait être en effet reconnu que comme une structure unique.

Chacune des structures nuageuses détectées peut donc être considérée comme étant un « objet » compact, objet auquel on peut par conséquent attribuer une dimension verticale (épaisseur

de la structure nuageuse) et horizontale (longueur de la structure nuageuse) (voir la Figure 2.5.1 (b)). Un algorithme de reconnaissance de forme a donc été mis en place, à partir des données LITE.

Après analyse des statistiques obtenues, nous avons aussi jugé nécessaire d'ajouter à cette première étape de reconnaissance de forme, une étape permettant l'agglomération des structures faiblement éloignées les unes des autres.

Cette étape d'agglomération est particulièrement justifiée lorsque nous abordons l'analyse des nuages semi transparents, tel que les cirrus, ou les résultats, après seuillage par notre algorithme de détection, montrent une structure horizontale géométriquement très fractionnée. Les zones entourées en pointillés rouges dans la Figure 2.4.7 sont les exemples de telles situations.

A cette étape de notre algorithme, ce résultat nous amènerait à déterminer un trop grand nombre de structures nuageuses de faible dimension, alors que les cirrus sont connus pour être des structures nuageuses de grande dimension, s'étendant sur des dizaines, voir des centaines de kilomètres. Dans les mêmes conditions, l'œil humain par exemple, ou encore un détecteur dont la résolution horizontale serait moindre que celle de LITE verrait automatiquement une structure compacte s'étendant sur de grandes échelles spatiales.

Un autre exemple où l'agglomération des structures est jugée utile concerne l'analyse de la base des structures nuageuses. Dans le cas de nuages absorbants, ou bien encore dans le cas de nuages recouverts, l'énergie chute alors rapidement lors de la traversée de la structure, et le signal rétrodiffusé n'est plus alors constitué que de bruit. Nous nous retrouvons alors rapidement à la limite des capacités de notre algorithme de détection à seuil, et les bases de ces structures se retrouvent alors fortement morcelées. Ces dernières structures, généralement de petites tailles, ne doivent pas être identifiées comme des structures isolées, mais doivent plutôt être rattachées à la structure nuageuse principale à laquelle elles appartiennent.

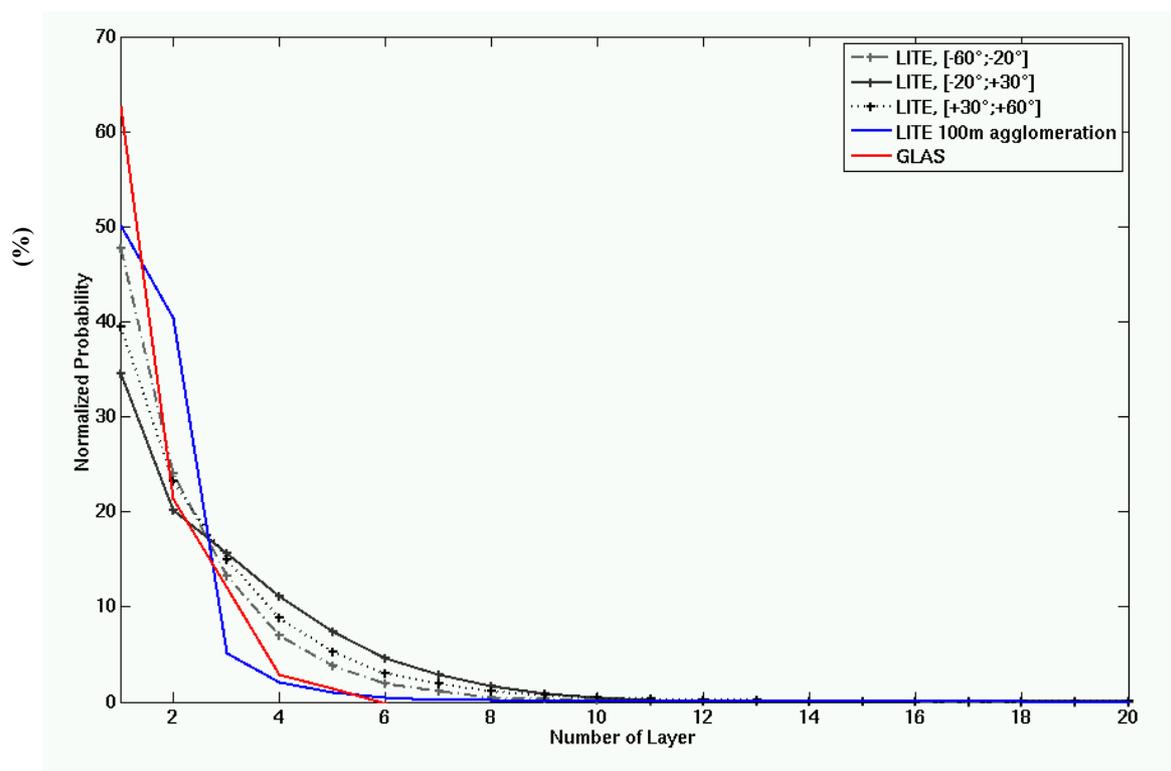
Nous détaillons dans la section suivante la méthode utilisée pour effectuer cette agglomération ainsi que les paramètres choisis.

## 2.5.2 La méthode d'agglomération

La difficulté de la mise en place d'un tel algorithme d'agglomération réside dans la détermination des distances limites sur la verticale ( $E_v$ ) et sur l'horizontale ( $E_h$ ) (voir la *Figure 2.5.1(b)*) pouvant séparer deux structures nuageuses, et qui amènerait ces deux structures à n'en constituer qu'une seule. Une étude de cas a donc été menée. Ces valeurs de  $E_v$  et de  $E_h$  sont fortement dépendantes de la résolution de l'instrument utilisé.

La distribution du nombre de couches nuageuses détectées sur la verticale constitue un paramètre d'importance permettant de juger de la qualité des agglomérations, et ainsi de celle des climatologies obtenues à partir de GLAS et de LITE.

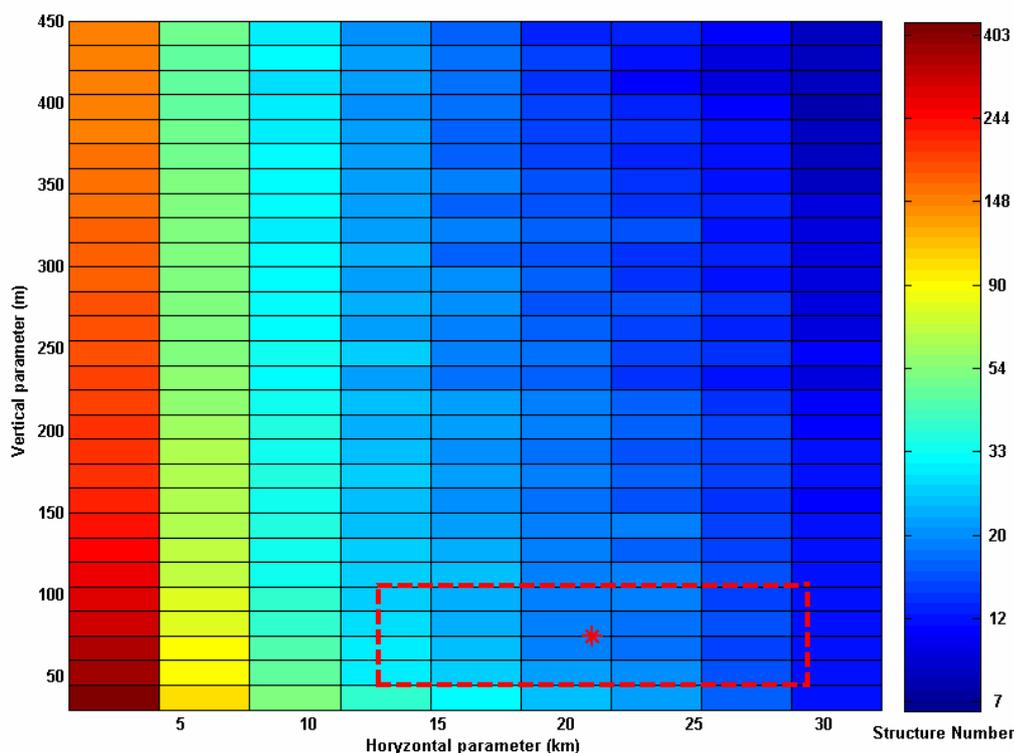
En effet, le nombre de couches nuageuses détectées sur la verticale par un lidar spatial doit directement être lié à sa résolution verticale. Or nous savons que la résolution verticale de LITE (15 m) est bien supérieure à celle de GLAS (76.8 m). Il semble donc intéressant de comparer les distributions du nombre de couches nuageuses correspondantes, et de juger de leur dépendance en fonction de la résolution verticale des instruments (voir la *Figure 2.5.2*).



**Figure 2.5.2 :** Distributions du nombre de couches détectée à partir des données LITE et des données GLAS. Les données de LITE à pleine résolution (15 m) sont étudiées suivant les trois bandes de latitude. La courbe bleue donne la distribution retrouvée après agglomération des structures dont la proximité est inférieure à 100 mètres. La courbe rouge donne la distribution du nombre de couches correspondant à GLAS.

La **Figure 2.5.2** donne les distributions du nombre de couches nuageuses de LITE à pleine résolution suivant les trois bandes de latitudes. On observe tout d'abord que le nombre de couches observées sur la verticale ne semble pas être fortement lié à la zone de latitude. Nous pouvons ensuite comparer ces distributions avec la distribution correspondante de GLAS. On remarque que le nombre de couches détectées par GLAS semble être bien moins important que dans le cas de LITE.

Afin de vérifier l'hypothèse de dépendance à la résolution, nous nous sommes proposés de retrouver quelle serait la distribution de LITE pour une résolution dégradée. Afin de remplir cet objectif, nous « agglomérons » les structures éloignées l'une de l'autre d'une distance maximum  $L$ , prise équivalente ici à 100 m. On observe alors que la proportion des tirs montrant plus de 4 structures chute alors rapidement, au profit des tirs monocouches et bicouches représentant alors l'essentiel de la statistique. La distribution de GLAS, de résolution verticale on le rappelle 76.8 m, semble bien alors constituer une distribution intermédiaire par rapport aux distributions de LITE, de résolution verticale 15 m à pleine résolution et celle agglomérée à 100 m.



**Figure 2.5.3** : Occurrence du nombre de structures nuageuses observées, en fonction des paramètres d'agglomération horizontaux  $E_h$  et verticaux  $E_v$

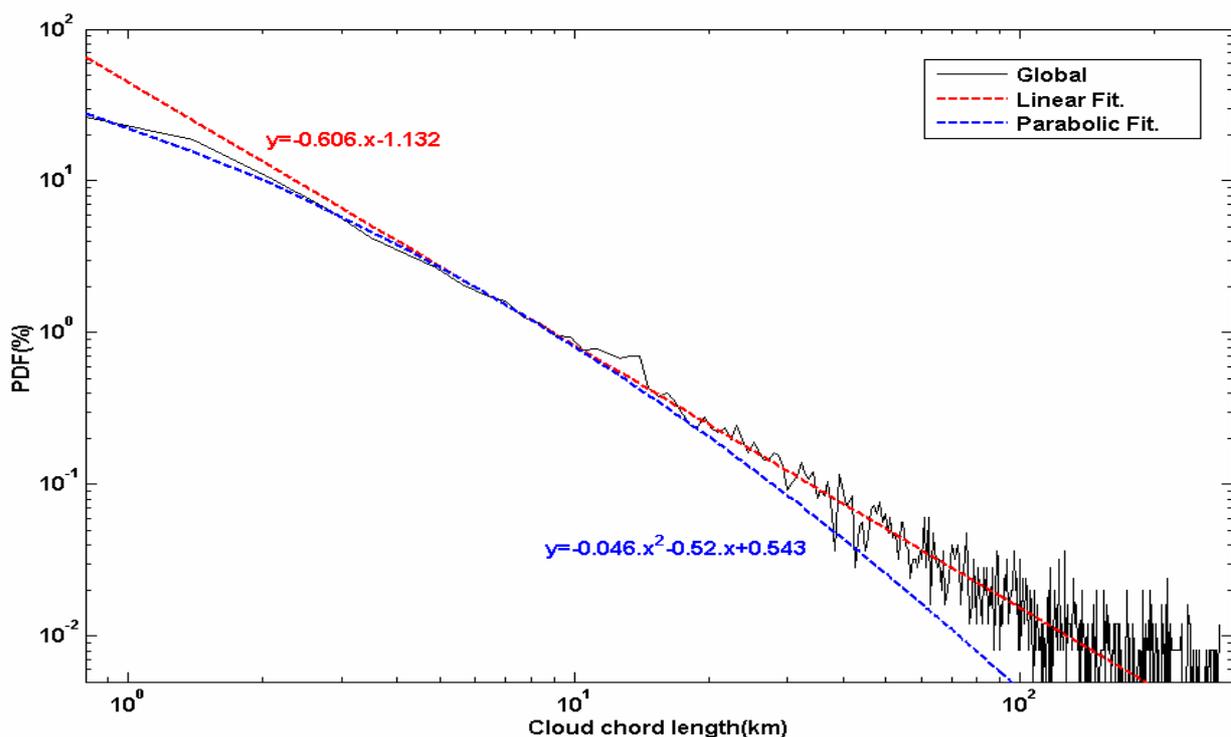
L'étude du nombre de structures nuageuses détectées (voir la **Figure 2.5.3**), a permis d'évaluer la qualité de l'agglomération des structures en fonction des distances limites d'agglomération, sur la verticale et sur l'horizontale. Les paramètres optimaux qui ont ainsi pu être déterminés correspondent à une distance limite d'agglomération sur l'horizontale  $E_h$  de 21 km (soit 30 pixels), et une distance limite d'agglomération sur la verticale  $E_v$  de 75 mètres (soit 5 pixels). Ces valeurs de  $E_v$  et de  $E_h$  correspondent à des échelles spatiales compatibles avec celles des modèles méso-échelles.

### 2.5.3 Résultats et comparaison

L'étape de reconnaissance des formes nous a permis de relier l'information contenue par chacun des tirs lidar, et par conséquent d'accéder à la distribution horizontale des nuages le long de l'orbite de LITE. La **Figure 2.5.4** présente la distribution de la longueur de structures telle que retrouvée après application de notre algorithme de reconnaissance de forme. Cette PDF (Probability Density

Function) montre une plus forte occurrence des structures de faibles longueurs que des structures de grandes longueurs. Les fortes variations autour de cette distribution pour les grandes longueurs de structures sont directement liées à leur faible nombre.

Cette distribution peut être approximée, en échelle logarithmique, pour les structures d'une longueur supérieure à 8 km par un fit linéaire  $y=-0.606x-1.132$ , et en deçà de 8km par un fit parabolique d'équation  $y= -0.046x^2-0.52x+0.543$ .



**Figure 2.5.4 :** Distribution de la longueur des structures nuageuses détectées à partir des données LITE

Il n'existe à ce jour que très peu d'études effectuées sur la longueur des structures nuageuses. A la différence des détecteurs passifs, l'instrument lidar est le seul à pouvoir offrir une estimation de la distribution de ces longueurs de structures par niveau d'altitude. Ces derniers apportent en effet une mesure intégrée sur la verticale, ce qui ne permet pas, dans le cas de système multicouche, d'avoir un accès direct à la longueur de chacune de ces structures nuageuses.

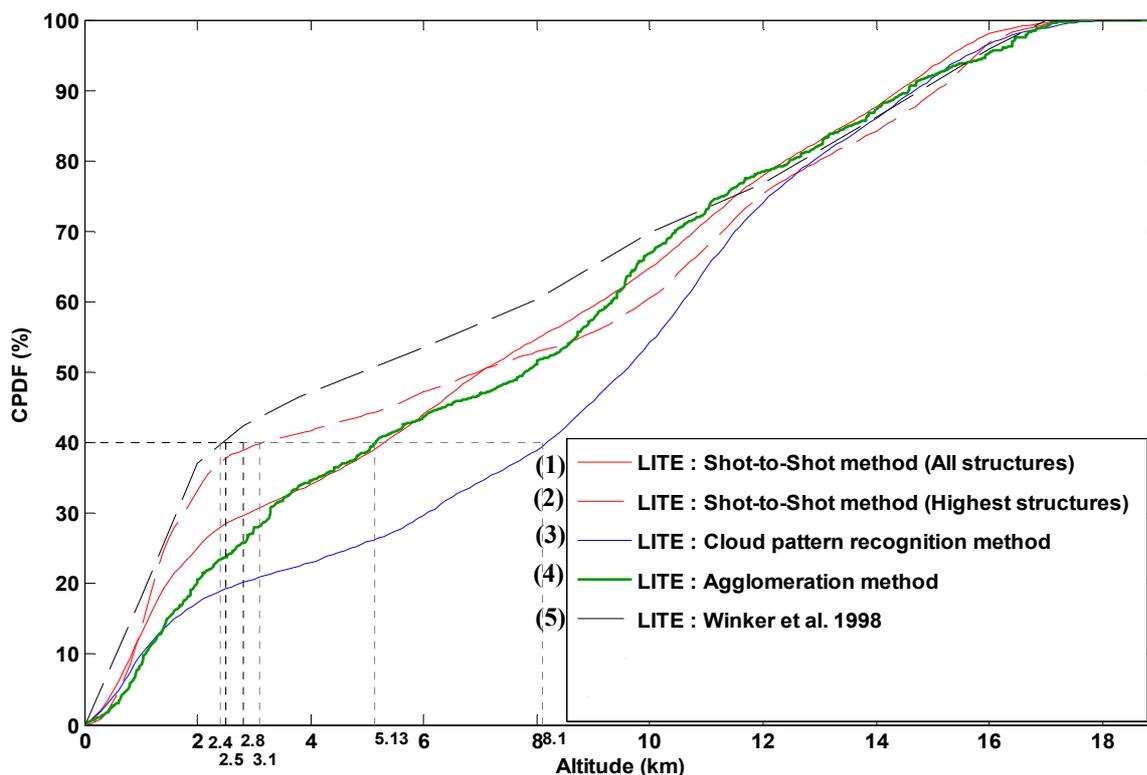
Des distributions de longueurs de structures obtenues à partir de stations in situ ont déjà été établies par d'autres auteurs, principalement dans le cadre des mesures effectuées par le réseau lidar terrestre américain ARM [Lane, D. E. et al, 2002]. Cependant, les analyses statistiques ainsi obtenues restent caractéristiques des régions où les mesures ont été effectuées, et restent difficilement comparables avec les résultats globaux apportés ici (différence d'échelle). De plus, la

détermination de la longueur des structures nuageuses, faite à partir d'une station lidar sol fixe ne peut s'effectuer de manière instantanée sur l'ensemble de la structure à caractériser. Elle constitue une méthode de mesure indirecte, car elle nécessite la connaissance des champs de vent à l'altitude correspondant à la structure nuageuse observée.

D'autres auteurs ont déjà fait l'analyse statistique de la distribution des longueurs de structures nuageuses, en employant par exemple les données du Landsat Multispectral Scanner (MSS), ou celles du Thematic Mapper (TM) (e.g. [Wielicki et Welch, 1986]; [Cahalan et Joseph, 1989]; [Davis et al., 1997]). [Wielicki et Welch, 1986] ont analysé la distribution nuageuse à partir des images de MSS, et ont montré que la distribution des longueurs de structures nuageuses retrouvées était comparable avec celle retrouvée dans les études précédentes basées sur l'étude des photographies aériennes ([Plank, 1969] et [Hozumi et al., 1982]), et celle des données TM ([Cahalan and Joseph, 1989]). Ils ont ainsi retrouvé que la distribution des longueurs était fonction du logarithme de la racine de la surface de la structure nuageuse, où le diamètre effectif de la structure nuageuse est approximé par une double loi de puissance, avec une puissance  $-0.89$  pour les structures de longueurs inférieures à 2 km, et  $-2.76$  pour les structures de tailles plus importantes.

Ces estimations ne sont pas concordantes a priori avec celles obtenues à partir des données LITE. On pourrait en effet adopter pour notre distribution le même type de loi de puissance, mais avec une puissance d'environ  $-0.52$  pour les structures de longueurs inférieures à 3 km, et  $-0.606$  pour les structures de tailles plus importantes.

La distribution globale des longueurs de structures nuageuses obtenues grâce aux données du lidar spatiale LITE constitue donc un résultat unique. Les résultats établis ici restent néanmoins difficilement comparables avec ceux précédemment établis par les stations in situ, et ne pourront être réellement comparés et vérifiés que grâce aux résultats qui pourront être obtenus à partir des données du lidar spatial CALIOP/CALIPSO.



**Figure 2.5.5 :** CPDF de l'altitude des sommets des structures nuageuses obtenus selon différentes méthodes : selon le mode tir à tir ( **(1)** en considérant l'ensemble des structures nuageuses, ou **(2)** seulement la structure la plus élevée de chaque tir), ou en regardant l'information sur la dimension horizontale : **(3)** après avoir effectué l'étape de reconnaissance de forme sur les structures nuageuses (Cloud pattern recognition method), et **(4)** après agglomération des structures précédemment reconnues (Agglomeration method). La CPDF **(5)** correspondante aux extractions effectuées en mode tir à tir par Winker et al. [1998], est aussi donnée dans cette figure.

Comme il a été effectué précédemment avant l'application du processus d'agglomération, nous pouvons maintenant fournir les nouvelles statistiques correspondantes aux sommets de structures nuageuses. La **Figure 2.5.5** présente la CPDF du sommet des structures nuageuses détectées. Nous avons fourni à titre de comparaison les CPDF correspondant à chacune des étapes menant à l'agglomération des structures nuageuses.

Nous remarquons tout d'abord de fortes différences entre la CPDF correspondante au mode d'analyse tir à tir (structure la plus haute sur chaque tir (Highest structure)), et celle correspondante au mode d'analyse après reconnaissance des formes. On observe une baisse du nombre de structures de moyenne altitude (classe M), au profit des structures de plus haute altitude. Ces sommets de structures (M) disparues correspondaient en fait à des trous au sein des structures nuageuses. Ces trous généraient alors au sein des statistiques des structures nuageuses « fantômes », biaisant par

conséquent ces dernières. L'apport de l'information sur la dimension horizontale a donc permis de s'affranchir de ces biais.

Après l'agglomération des structures nuageuses nouvellement déterminées, et selon les paramètres d'agglomération  $E_v$  et  $E_h$  tel que déterminés par l'étude de sensibilité effectuée dans la section précédente, nous présentons maintenant la nouvelle CPDF correspondante. Cette dernière montre pour sa part une évolution plus régulière sur l'ensemble de l'altitude.

Nous nous intéressons maintenant à la **Table 2.8**, qui présente la nouvelle statistique nuageuse établie à partir des données LITE, et après agglomération des structures nuageuses. Cette statistique peut être directement comparée avec celle fournie avant agglomération par la **Table 2.4**.

Highest		All latitude range	[-60°;-20°]	[-20°;+30°]	[+30°;+60°]
High	H	58.8 (52.9 ; 67.8) % 58.0 (51.1 ; 67.1) %	49.9 (48.2 ; 64.7) % 51.9 (50.4 ; 63.9) %	60.8 (56.0 ; 67.5)% 59.5 (49.4 ; 68.9)%	60.2 (51.3 ; 68.3) % 59.0 (54.4 ; 64.2) %
	H+M	13.1 (11.4 ; 15.4) % 12.6 (10.5 ; 15.3) %	13.0 (11.7 ; 22.3) % 12.3 (10.1 ; 24.6) %	11.5 (9.9 ; 13.8) % 11.4 (8.8 ; 13.8) %	15.5 (14.0 ; 17.0) % 15.2 (14.1 ; 16.4) %
	H+L	25.0 (32.0 ; 14.2) % 26.3 (34.8 ; 15.2) %	33.1 (35.7 ; 12.2) % 32.3 (35.5 ; 11.0) %	24.8 (31.3 ; 15.7)% 26.2 (38.6 ; 14.7)%	21.4 (30.2 ; 12.4) % 22.3 (27.2 ; 17.0) %
	L+M+H	3.1 (3.7 ; 2.6) % 3.1 (3.6 ; 2.4) %	4.0 (4.4 ; 0.8) % 3.5 (4.0 ; 0.5) %	2.9 (2.8 ; 3.0) % 2.9 (3.2 ; 2.6) %	2.9 (4.5 ; 2.3) % 3.5 (4.3 ; 2.4) %
Middle	M	75.3 (68.3 ; 86.8) % 76.1 (67.8 ; 87.8) %	64.1 (62.0 ; 92.8) % 65.4 (63.6 ; 92.6) %	77.4 (65.7 ; 86.0) % 78.4 (62.9 ; 90.5) %	81.4 (75.9 ; 85.8) % 81.1 (76.3 ; 85.5) %
	M+L	24.7 (31.7 ; 13.2) % 23.9 (32.2 ; 12.2) %	35.9 (38.0 ; 7.2) % 34.6 (36.4 ; 7.4) %	22.6 (34.3 ; 14.0) % 21.6 (37.1 ; 9.5) %	18.6 (24.1 ; 14.2) % 18.9 (23.7 ; 14.5) %

**Table 2.8** : Répartitions nuageuses dans chaque classe d'altitude de LITE sous la couverture ISCCP-D1 (chiffres noirs) et ISCCP-DX (chiffres rouges), après agglomération. Le premier chiffre correspond à la couverture globale. Le second chiffre correspond à la couverture au dessus de la mer, et le troisième au dessus de la terre. Pour chacune des classes ISCCP, nous avons reporté le pourcentage des classes monocouches multicouches.

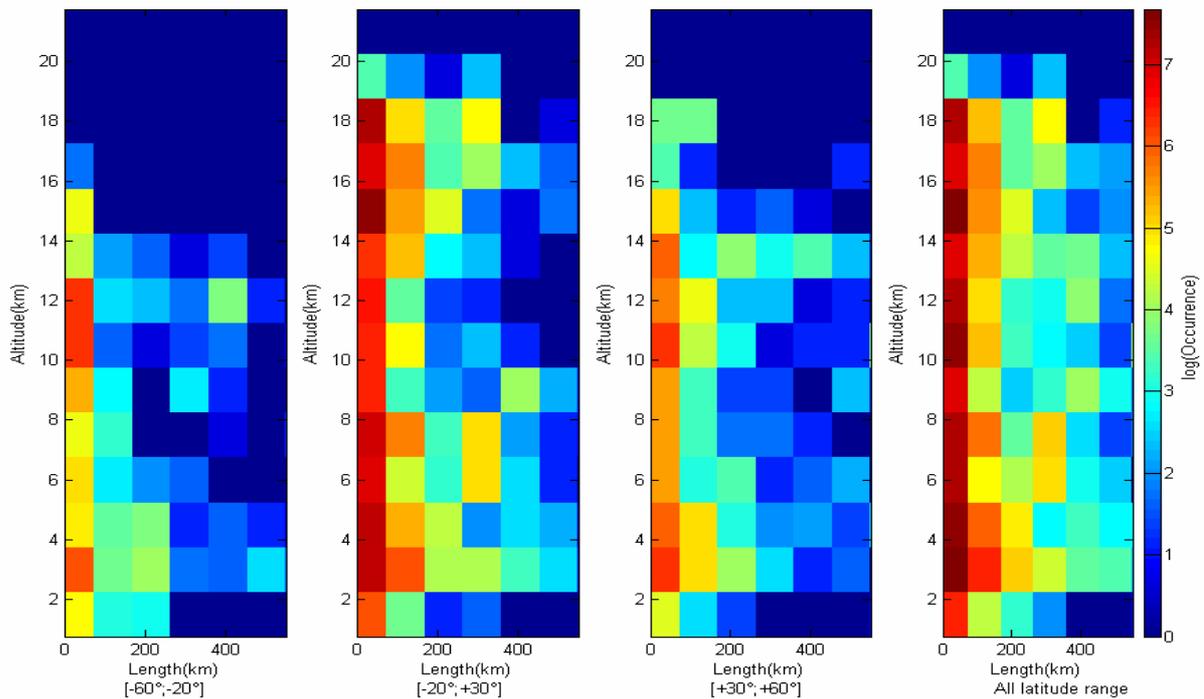
On note tout d'abord, comme ce fut le cas avant l'étape d'agglomération des structures nuageuses, que les statistiques nuageuses sont différentes suivant les bandes de latitudes considérées.

De plus, on assiste, comparativement à la climatologie présentée avant agglomération (voir la **Table 2.4**), à une augmentation générale du nombre de structures monocouches. Cet effet se comprend aisément par l'action de l'agglomération, qui tend à regrouper en une seule structure des structures précédemment fractionnées. Par exemple, les structures montrant de fortes extensions spatiales, comme dans le cas par exemple des cumulonimbus, se retrouvent maintenant solidarisées. Il s'effectue donc un rééquilibrage vers les hautes altitudes des statistiques avant agglomération. La **Table 2.9** montrant la statistique des structures les plus hautes sur chacun des tirs lidar (correspondant à la **Table 2.5** avant agglomération), montre une chute globale de l'occurrence des structures basses (on passe de 38.6 à 24.87%), au profit des structures hautes (de 52.4 à 54.47%), et surtout au profit des structures moyennes (de 9 à 20.64%). Ce rééquilibrage tend à confirmer celui observé précédemment dans les CPDF (voir la **Figure 2.5.5**).

Highest	All latitude range	[-60°;-20°]	[-20°;+30°]	[+30°;+60°]
<b>High</b>	54.47 % 55.05 %	54.47 % 55.05 %	51.55 % 49.76 %	51.55 % 49.76 %
<b>Middle</b>	20.64 % 13.18 %	9.44 % 7.06 %	24.82 % 13.31 %	14.49 % 19.02 %
<b>Low</b>	24.87 % 31.77 %	20.91 % 19.38 %	23.61 % 36.32 %	31.67 % 34.11 %

**Table 2.9** : Statistiques nuageuses des données LITE après agglomération. Les chiffres en noir correspondent à la couverture des données ISCCP-D1. Les chiffres en rouge correspondent à celle des données ISCCP-DX.

Grâce à l'apport de l'information sur la dimension verticale, nous avons pu étudier les deux paramètres climatologiques d'importance que constituent l'altitude des structures nuageuses, ainsi que la dimension de ces nuages. Nous allons maintenant chercher, pour clore cette étude des données actives, à faire le lien entre ces deux variables.



**Figure 2.5.6 :** Lien entre altitude du sommet et dimension du nuage (la taille des boîtes sur la longueur est de 50 km, pour une taille sur l'altitude de 1.5 km)

La **Figure 2.5.6** nous donne en global, et suivant les zones de latitude, le lien entre ces deux paramètres. Nous observons alors sur les trois zones d'altitudes, que l'occurrence des structures nuageuses ayant une grande longueur est importante dans le cas des nuages bas, et dans le cas des nuages de très haute altitude (entre 12 et 14 km d'altitude pour les zones tropicales, et entre 14 et 19 km d'altitude dans le cas de la zone intertropicale).

Cette observation est particulièrement vérifiée par l'observation dans le cas des nuages cumuliformes, qui apparaissent alors, après agglomération, avec des extensions horizontales très importantes (voir la **Figure 2.5.7**). De même, la longueur des structures de types cirrifformes, situées aux hautes altitudes telles que décrites dans le paragraphe précédent, correspond aux observations que l'on peut faire de ce type de nuages (voir la **Figure 2.5.7**).



**Figure 2.5.7:** Champ de Cumulus (image de gauche), et de cirrus (image de droite), photographiés par avion.

## **2.6 Conclusion du chapitre sur la partie nuage détectée par instrumentation active**

La comparaison des données LITE et GLAS nous a permis d'aboutir à plusieurs conclusions importantes quant à l'utilisation des données lidar afin de construire une statistique nuageuse. On a, dans un premier temps, montré que bien que les instruments lidar de la mission LITE et de la mission GLAS sont différents dans leurs caractéristiques, les statistiques nuageuses retrouvées restent cohérentes entre elles. Ce travail constitue une première étape de validation de notre algorithme d'extraction nuageuse, en vue du traitement des données de la mission CALIPSO.

Les différences caractérisant ces deux missions (durées cumulées d'enregistrement, homogénéité géographique des enregistrements, résolution verticale, caractéristiques instrumentales, etc ...), nous ont permis d'expliquer une grande part des différences observées entre les statistiques.

Dans le cadre de la mission CALIPSO, il sera intéressant de voir si cette cohérence des statistiques nuageuses sera conservée. L'étude des différences de conceptions caractérisant les trois

instruments LITE, GLAS et CALIOP, ainsi que l'étude de l'impact de ces différences sur les climatologies nuageuses retrouvées, permettront à terme de nous rapprocher au plus près de la réalité physique du couvert nuageux. Ce travail permettra en outre une extrapolation plus fine des performances potentielles des futurs lidar spatiaux.

Les études qui ont été produites dans ce chapitre nous ont permis de tirer parti des avantages qu'offre la technologie lidar embarquée sur un satellite. Cette technologie est en effet aujourd'hui la seule à pouvoir nous faire accéder à l'information, concernant la répartition sur la dimension verticale des structures nuageuses multicouches. Les extractions des données LITE et GLAS nous ont permis de dresser des statistiques nuageuses fiables, qui devraient être prises en compte dans les modèles de climatologies.

Nous avons aussi proposé dans ce chapitre une vision novatrice de la structure nuageuse détectée par les lidar, en dépassant le simple échantillonnage 1D (vision tir à tir). En profitant de l'information que nous avons sur la dimension horizontale, nous avons pu identifier les nuages en tant qu'objet compact, et ainsi permettre une restitution 2D des structures nuageuses.

Cette nouvelle connaissance de la distribution 2D des structures nuageuses, obtenue grâce à la télédétection active, constitue une des informations essentielles permettant, via le renseignement d'un modèle radiatif, d'estimer l'impact radiatif lié à la couverture nuageuse.