

Nuages de mi-niveau en Afrique de l'Ouest origine, propriétés et impacts

Elsa BOURGEOIS

Dominique Bouniol, Fleur Couvreur et Françoise Guichard

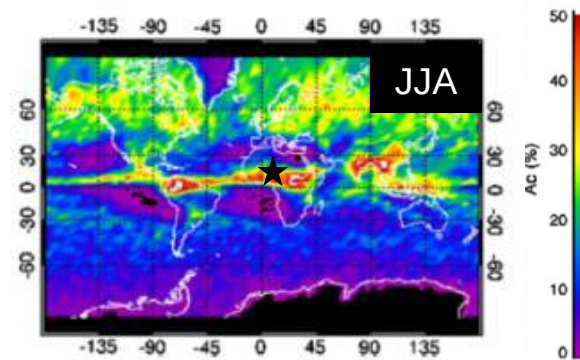
Plan

- I. Objectifs et contexte
- II. Propriétés des nuages de mi-niveau en Afrique de l'Ouest
 - a) Propriétés macrophysiques et microphysiques
 - b) Différentes familles de nuages
- III. Processus à l'origine de ces nuages de mi-niveau
 - a) Conditions thermodynamiques
 - b) Simulation 1D
- IV. Conclusions et perspectives

Où observe-t-on les nuages de mi-niveau ?

A l'échelle du globe

Abondance des nuages de mi-niveau
dans la ceinture tropicale.

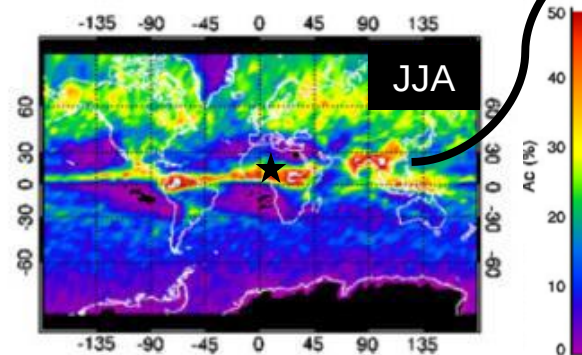


*Distribution globale des altocumulus. CloudSat
et CALIPSO. 2007 – 2008. (Sassen and Wang, 2011)*

Où observe-t-on les nuages de mi-niveau ?

A l'échelle du globe

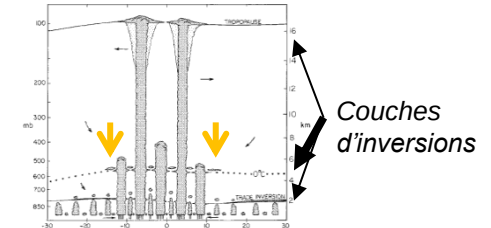
Abondance des nuages de mi-niveau dans la ceinture tropicale.



Distribution globale des altocumulus. CloudSat et CALIPSO. 2007 – 2008. (Sassen and Wang, 2011)

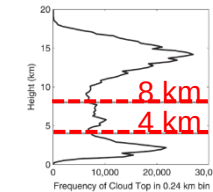
Dans le Pacifique Tropical Océanique

Distribution tri-modale des nuages.



Distribution verticale des nuages tropicaux : cumulus, cumulus congestus et cumulonimbus. Radar et radiosondages. (Johnson et al., 1999)

Station ARM à Darwin de 2005 à 2010.



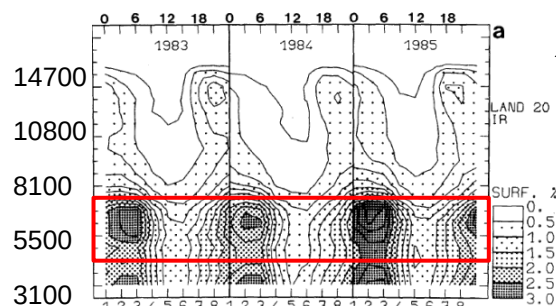
Nuages de mi-niveau

Hauteurs des sommets des nuages (Riihimäki et al., 2012)

Où observe-t-on les nuages de mi-niveau ?

En Afrique

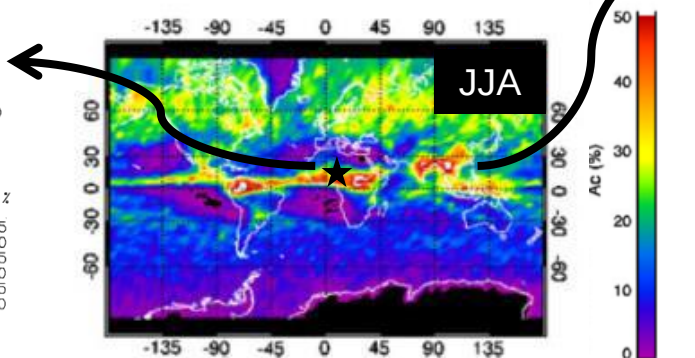
Première construction d'une climatologie de la nébulosité (observations spatiales - Météosat)



Evolution du cycle diurne en moyenne été 1983, 1984 et 1985. (Duvel, 1989)

A l'échelle du globe

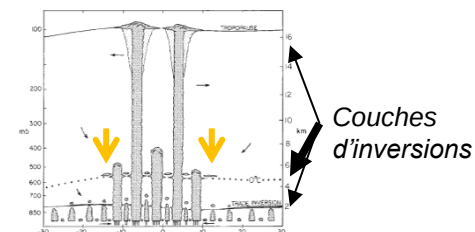
Abondance des nuages de mi-niveau dans la ceinture tropicale.



Distribution globale des altocumulus. CloudSat et CALIPSO. 2007 – 2008. (Sassen and Wang, 2011)

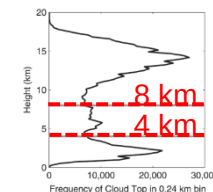
Dans le Pacifique Tropical Océanique

Distribution tri-modale des nuages.



Distribution verticale des nuages tropicaux : cumulus, cumulus congestus et cumulonimbus. Radar et radiosondages. (Johnson et al., 1999)

Station ARM à Darwin de 2005 à 2010.

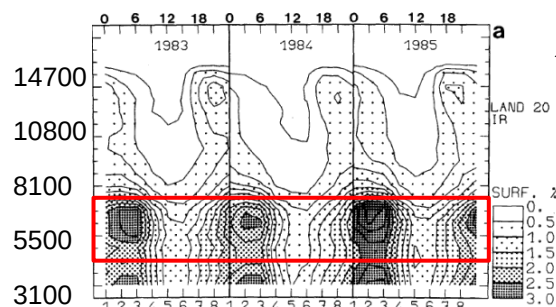


Hauteurs des sommets des nuages (Riihimäki et al., 2012)

Où observe-t-on les nuages de mi-niveau ?

En Afrique

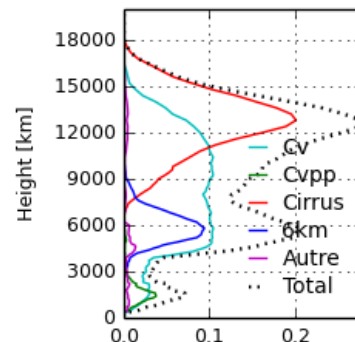
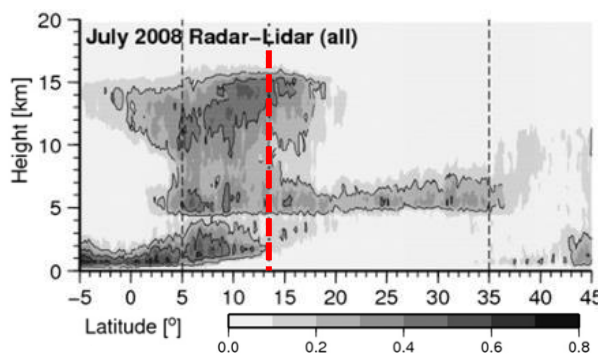
Première construction d'une climatologie de la nébulosité (observations spatiales - Météosat)



Evolution du cycle diurne en moyenne été 1983, 1984 et 1985. (Duvel, 1989)

En Afrique de l'Ouest / Station ARM à Niamey

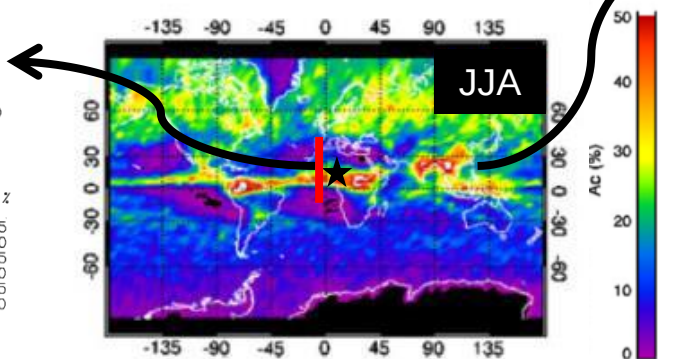
Fréquence d'occurrence des nuages. CloudSat-CALIPSO.



Fréquence d'occurrence d'avril à septembre des nuages à Niamey (ARM) (Bouniol et al., 2011)

A l'échelle du globe

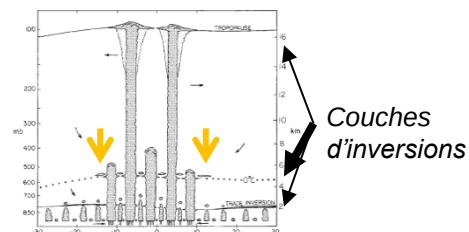
Abondance des nuages de mi-niveau dans la ceinture tropicale.



Distribution globale des altocumulus. CloudSat et CALIPSO. 2007 – 2008. (Sassen and Wang, 2011)

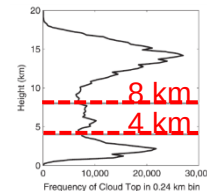
Dans le Pacifique Tropical Océanique

Distribution tri-modale des nuages.



Distribution verticale des nuages tropicaux : cumulus, cumulus congestus et cumulonimbus. Radar et radiosondages. (Johnson et al., 1999)

Station ARM à Darwin de 2005 à 2010.



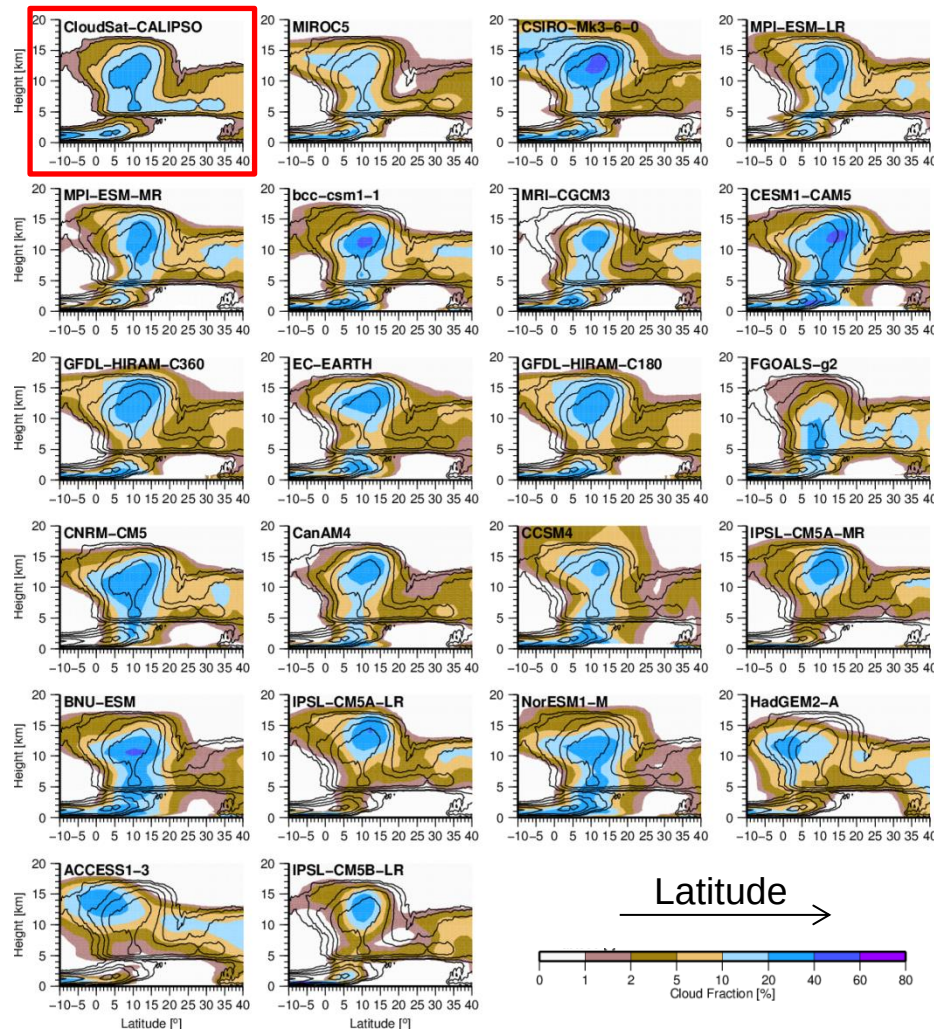
Hauteurs des sommets des nuages (Riihimaki et al., 2012)

Quelles sont les caractéristiques de ces nuages de mi-niveau en Afrique de l'Ouest ?

Quels sont les processus physiques à l'origine et permettant le maintien de ces nuages de mi-niveau en Afrique de l'Ouest ?

Quelle représentation des nuages de mi-niveau dans les modèles de climat ?

Incapacité des modèles de climat à reproduire l'occurrence des nuages de mi-niveau



Courbes latitude - altitude de la fraction nuageuse moyenne entre 10°W et 10°E pour les mois de juillet, août et septembre.

CloudSat-CALIPSO : 2006 - 2010

Modèles CMIP5 : 1979 - 2008

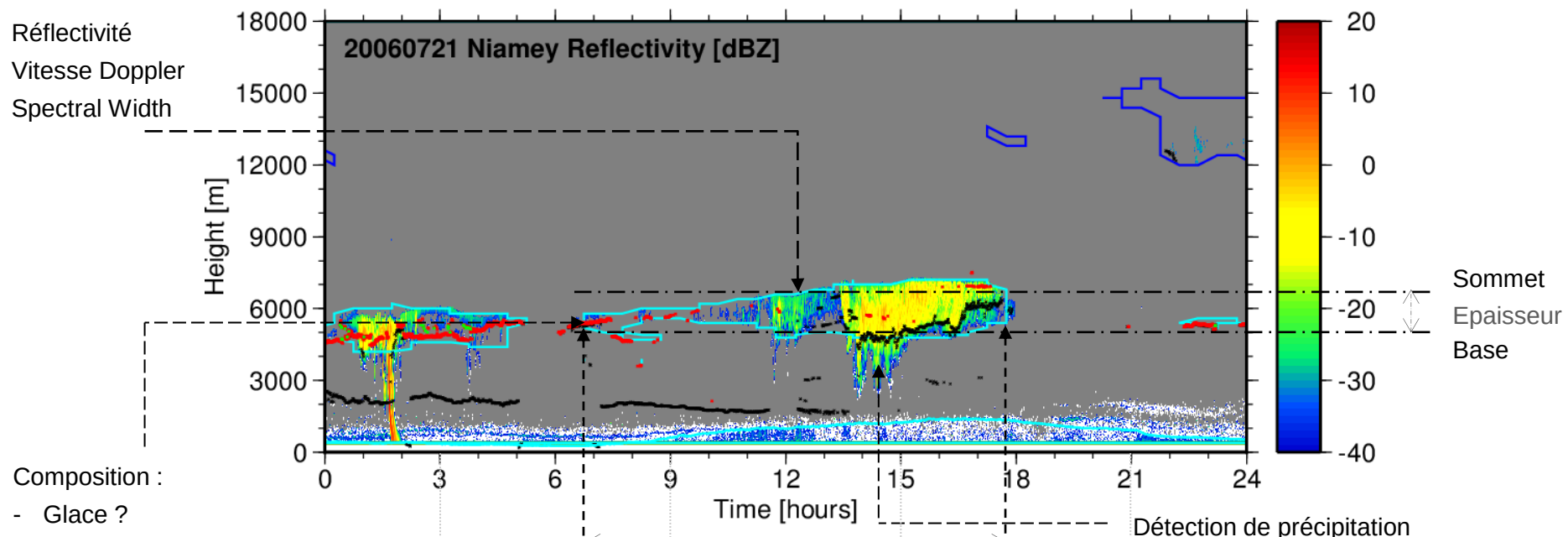
(Roehrig et al., 2013)



Pourquoi les modèles de climat ne simulent-ils pas correctement ces nuages de mi-niveau ?

Identification et caractéristiques des nuages de mi-niveau

Station ARM Niamey (Niger) - 2006



Composition :

- Glace ?
- Glace fondue ?
- Glace fondue et gouttelettes d'eau liquide ?
- Glace et gouttelettes d'eau liquide en surfusion ?
- Gouttelettes d'eau liquide ?
- Bruine ou pluie avec des gouttelettes d'eau liquide ?

Lidar / Radar

Δt 30sec

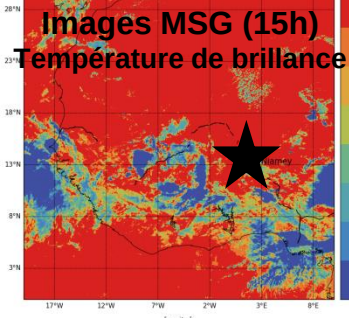
Δt 15min

Δt 6h

Date et heure
du début du nuage

Temps de
passage du
nuage

Date et heure
de la fin du nuage

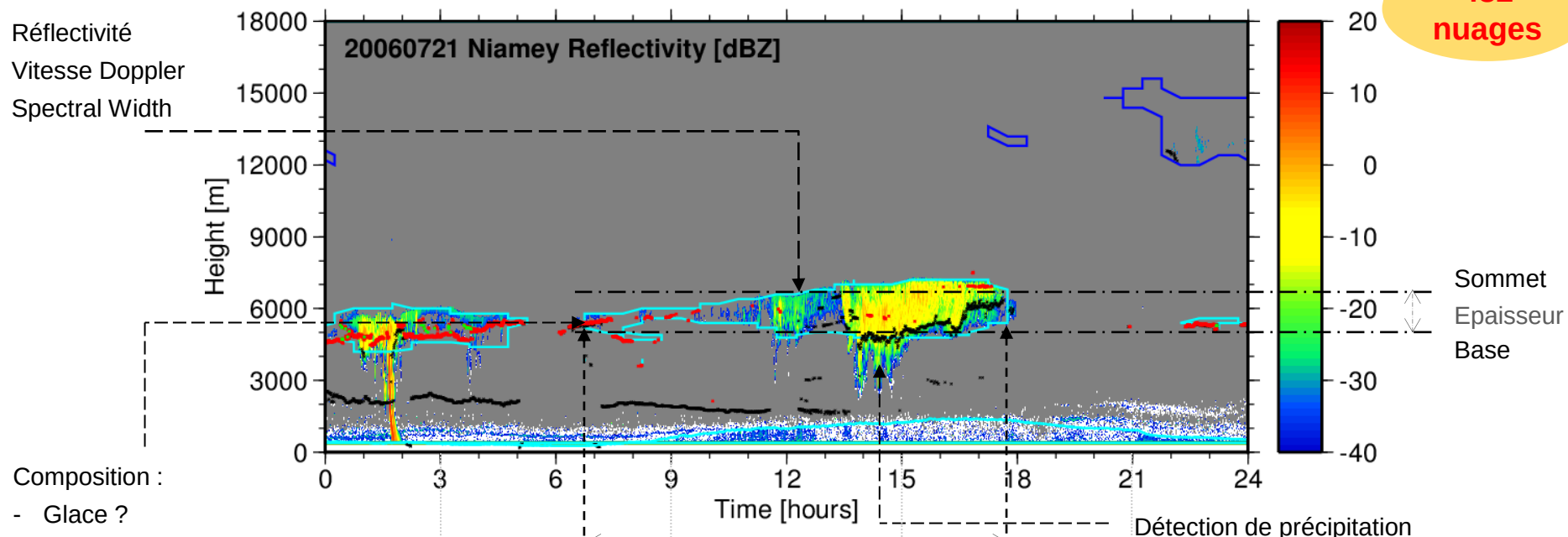


Images MSG

Radiosondages

Identification et caractéristiques des nuages de mi-niveau

Station ARM Niamey (Niger) - 2006



Composition :

- Glace ?
- Glace fondue ?
- Glace fondue et gouttelettes d'eau liquide ?
- Glace et gouttelettes d'eau liquide en surfusion ?
- Gouttelettes d'eau liquide ?
- Bruine ou pluie avec des gouttelettes d'eau liquide ?

Date et heure
du début du nuage

Temps de
passage du
nuage

Date et heure
de la fin du nuage

Lidar / Radar

Δt 30sec

Δt 15min

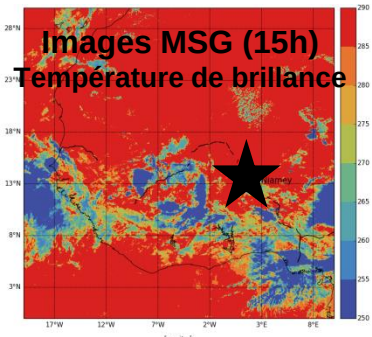
Δt 6h

4h45

5h32

17h34

19h45



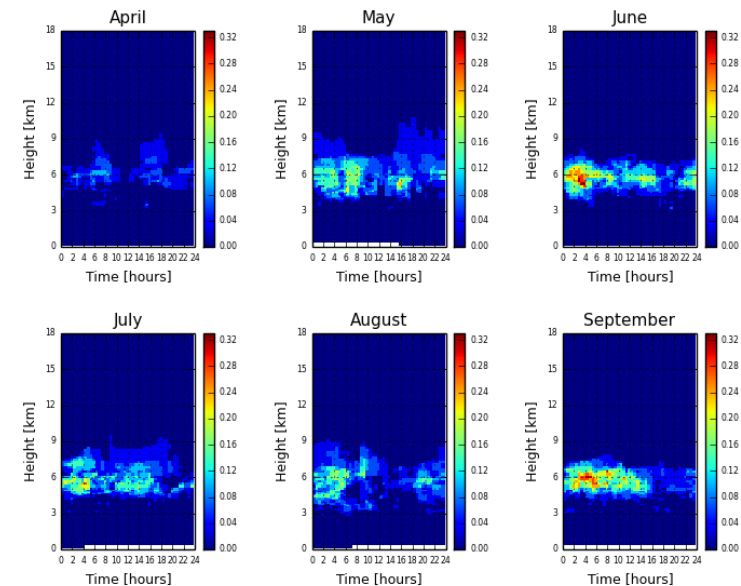
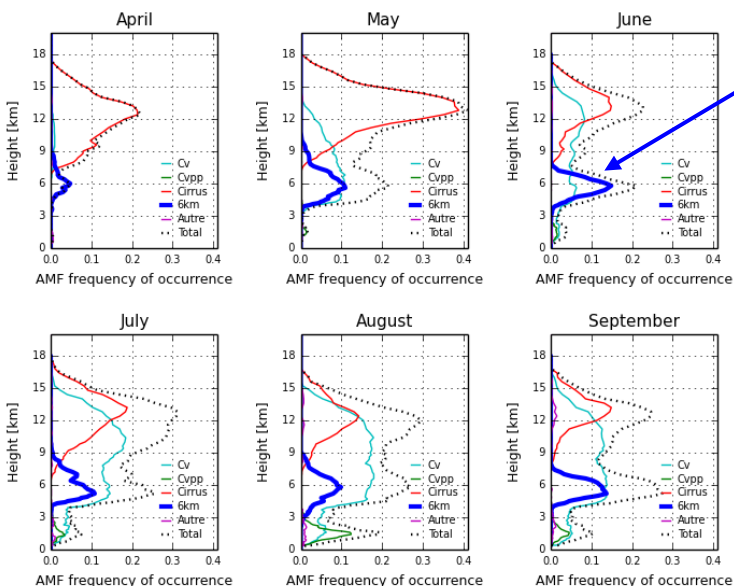
Images MSG

Radiosondages

Occurrence et cycle diurne des nuages de mi-niveau

Période de mousson

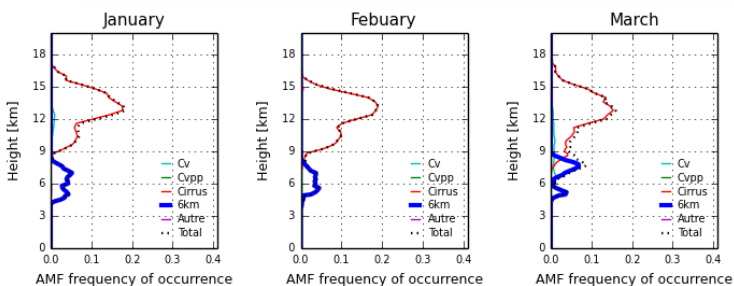
Occurrences des nuages (ARM Niamey 2006).



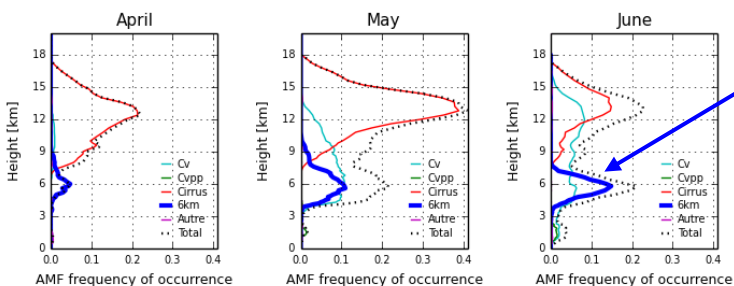
Cycle diurne mensuel des nuages de mi-niveau (ARM Niamey 2006).

Occurrence et cycle diurne des nuages de mi-niveau

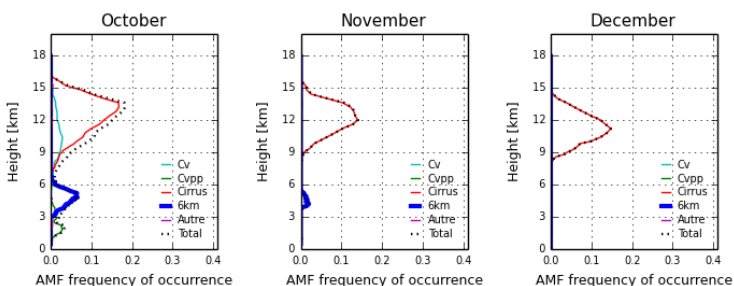
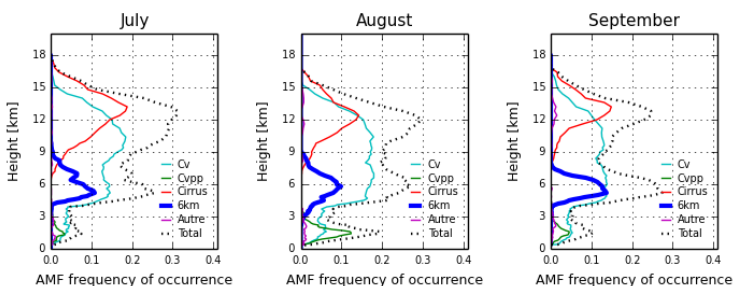
Année 2006



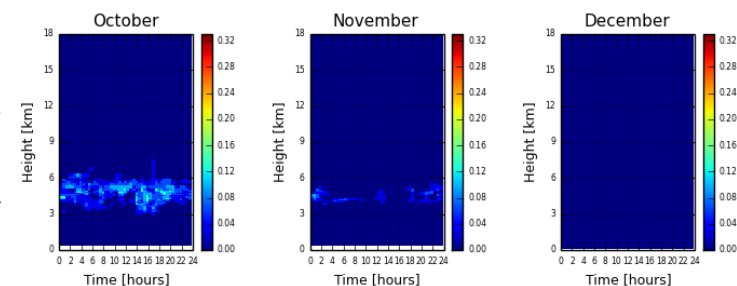
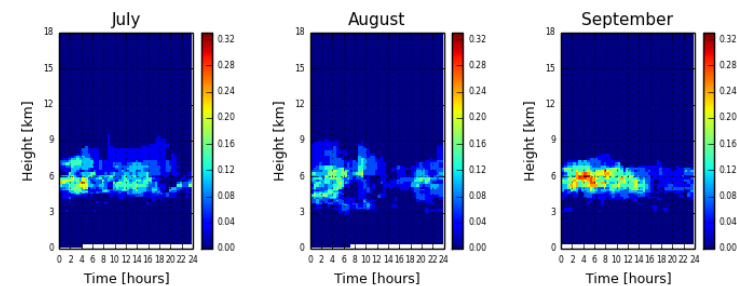
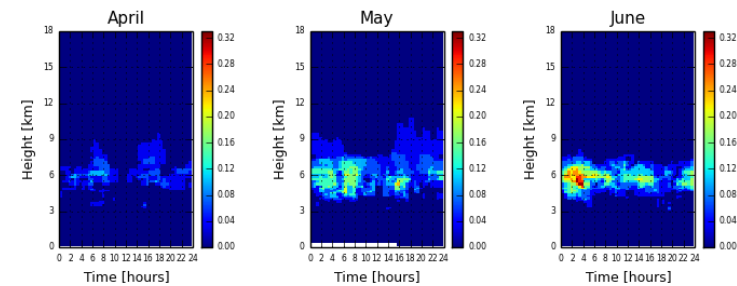
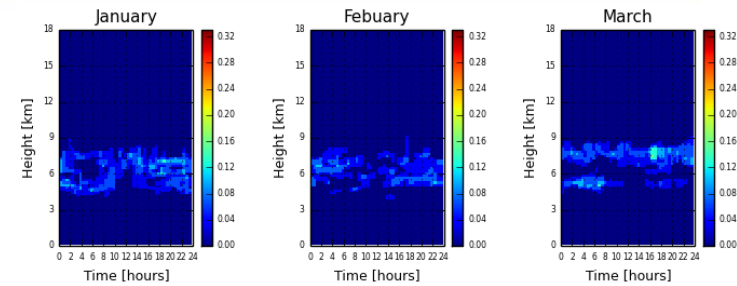
Occurrences des nuages (ARM Niamey 2006).



Nuages de mi-niveau



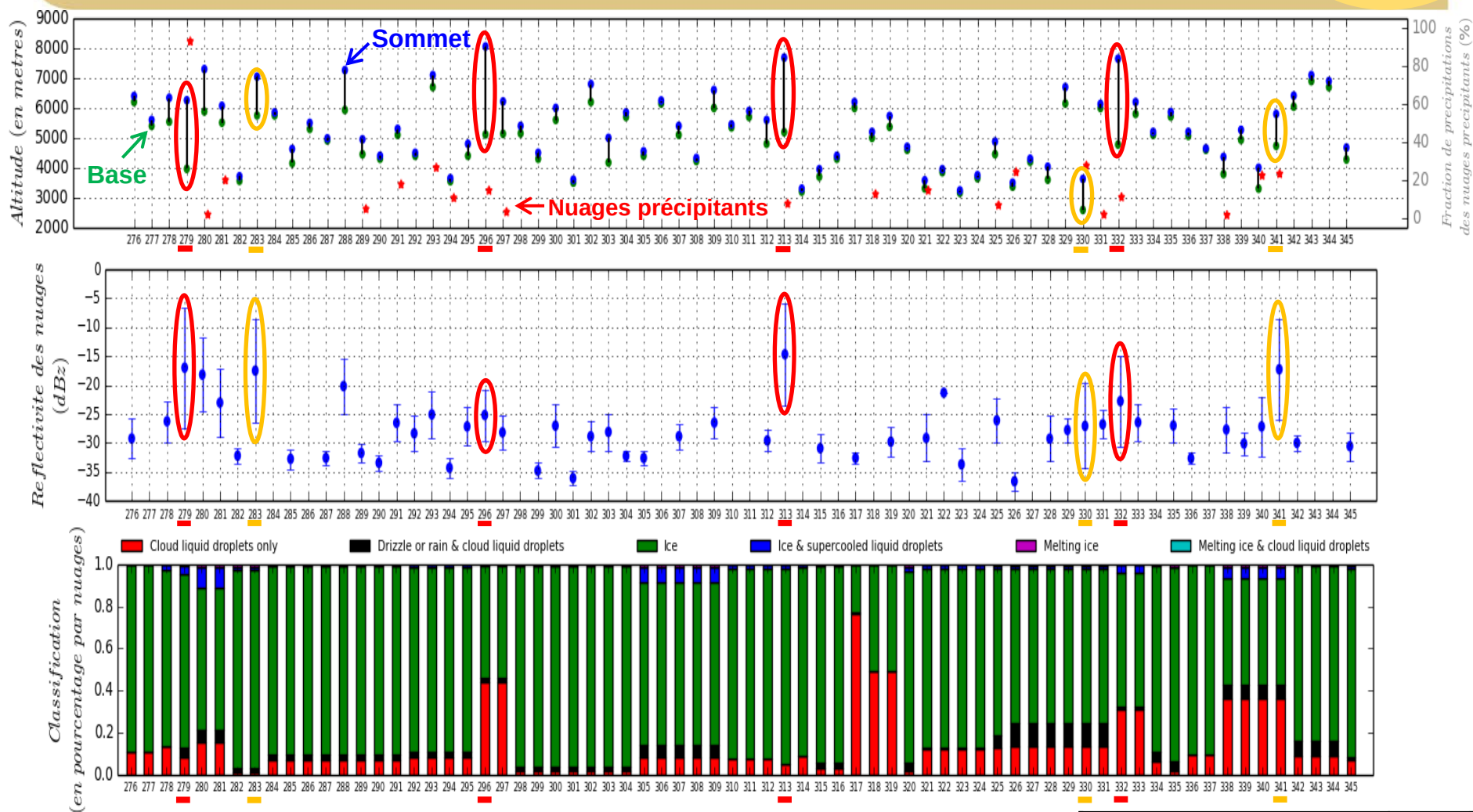
Cycle diurne mensuel des nuages de mi-niveau (ARM Niamey 2006).



Données et pré-traitements

Fiche d'identité du mois d'Août

482
nuages
en 2006



En janvier, février et mars

72

Détectés par le radar ($\geq 25\%$)

198

 ≤ 30 minutes

198

< 500 m d'épaisseurs

342

Hors janvier, février et mars

410

Non détectés par le radar ($< 25\%$)

284

> 30 minutes

284

 $500 \text{ m} \leq \text{épaisseurs} < 2000 \text{ m}$

130

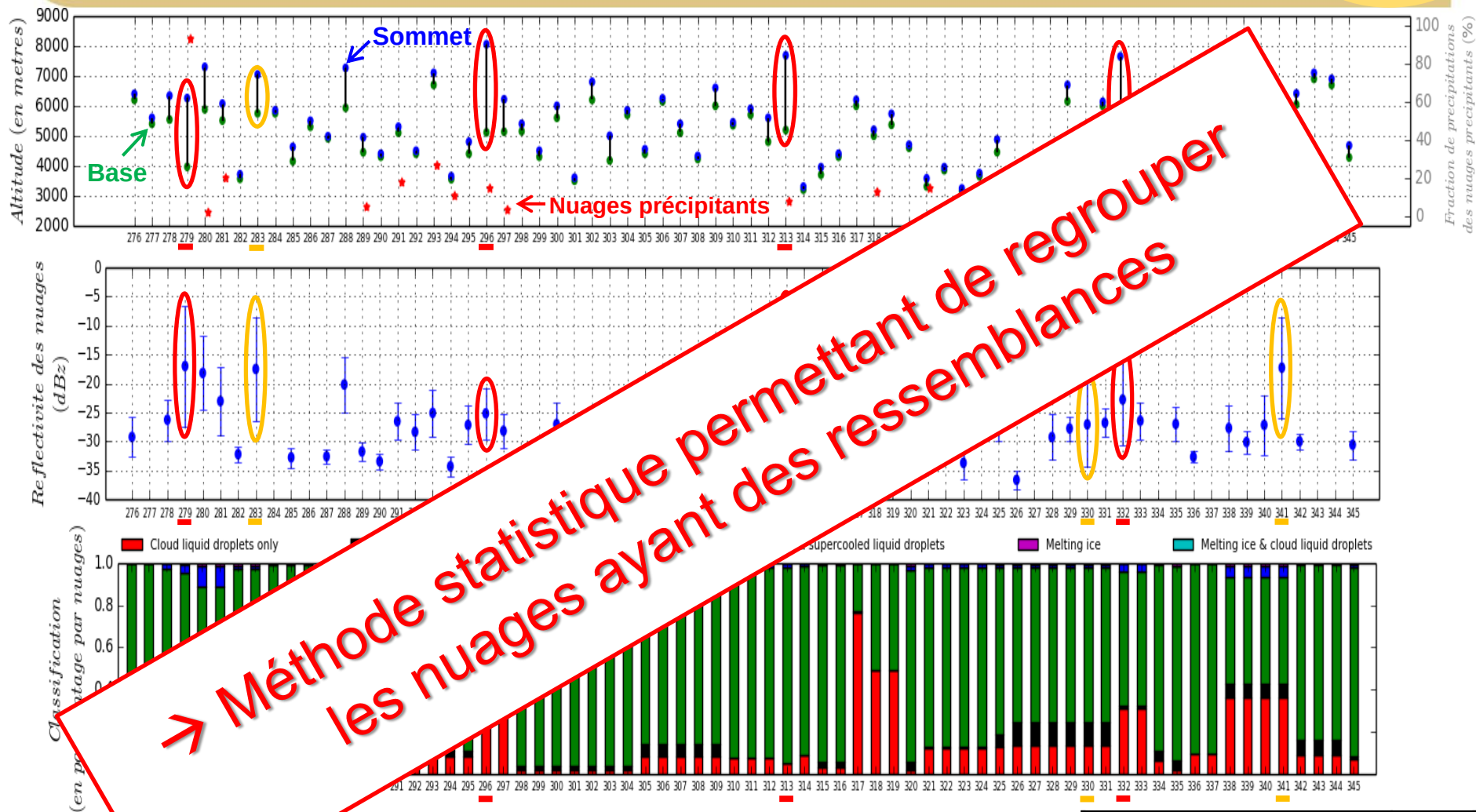
 $\geq 2000 \text{ m d'épaisseurs}$

10

Données et pré-traitements

Fiche d'identité du mois d'Août

482
nuages
en 2006



En janvier, février et mars	72
Hors janvier, février et mars	410

Détectés par le radar ($\geq 25\%$)	198
Non détectés par le radar ($< 25\%$)	284

≤ 30 minutes	198
> 30 minutes	284

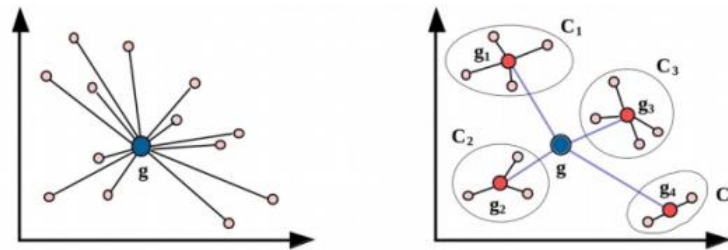
< 500 m d'épaisseurs	342
$500 \text{ m} \leq \text{épaisseurs} < 2000 \text{ m}$	130
≥ 2000 m d'épaisseurs	10

Méthodes de clustering

Qu'est ce que c'est ? - Les différentes méthodes

Objectif : Regrouper entre eux des objets similaires selon certains critères.

→ Grande similarité intra-groupe / Faible similarité inter-groupe.



Inertie totale des points = Inertie Intra-cluster + Inertie Inter-cluster

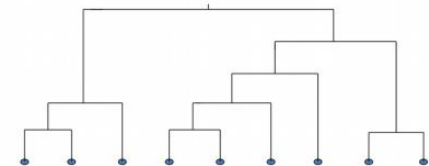
Deux types de méthodes :

– Méthodes de partitionnement (**non-déterministe**):

Diviser les objets en un nombre fixe de groupes (K). (**K-Means**)
Détermination de K en amont (méthode Elbow).

– Méthodes hiérarchiques (**déterministe**):

Produire une séquence de groupes imbriqués
les uns dans les autres → dendrogramme. (**Méthode de Ward**)



Dendrogramme

Tests :

Variables morphologiques (toute l'année 2006)

Bases et épaisseurs

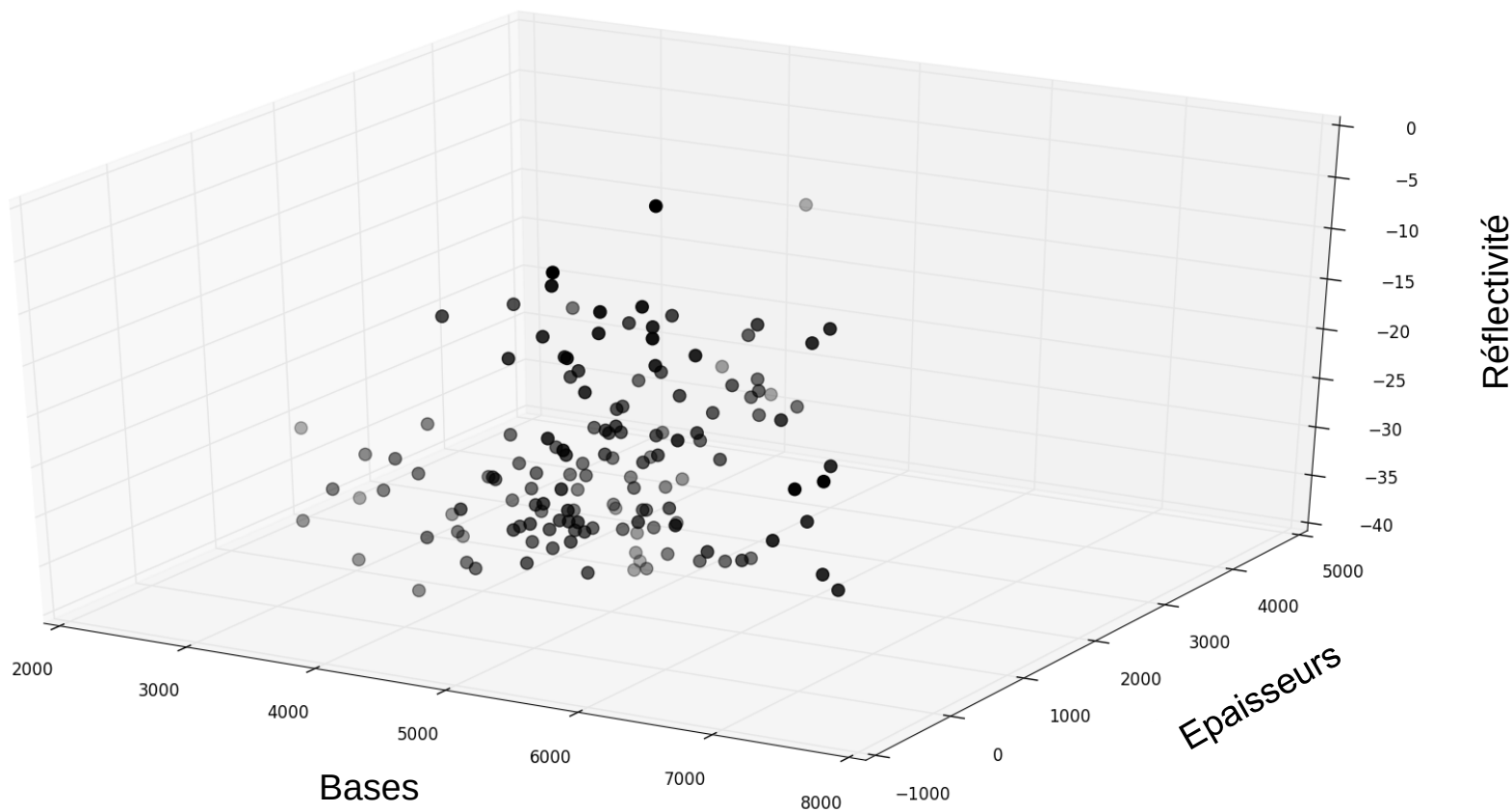
→ Résultats robustes et indépendants des méthodes

Toujours
résultats
avec ces
méthodes

Méthode de hiérarchisation Ward

Station ARM Niamey (Niger) - 2006

(pas JFM, durées > 30 min, détection radar $\geq 25\%$)

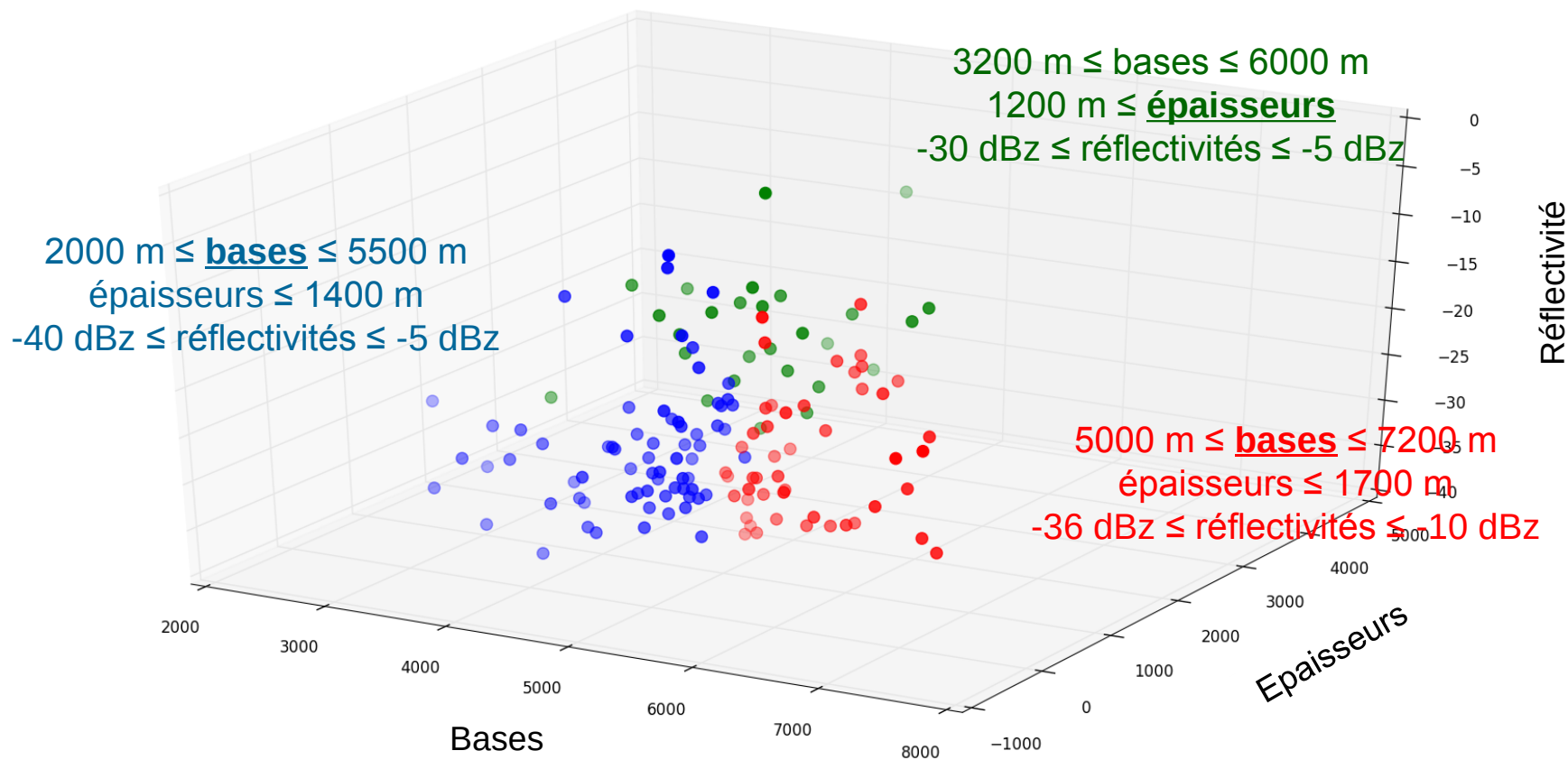


Utilisation des variables **bases** et **épaisseurs** de la base de données

Méthode de hiérarchisation Ward - Application

Station ARM Niamey (Niger) - 2006

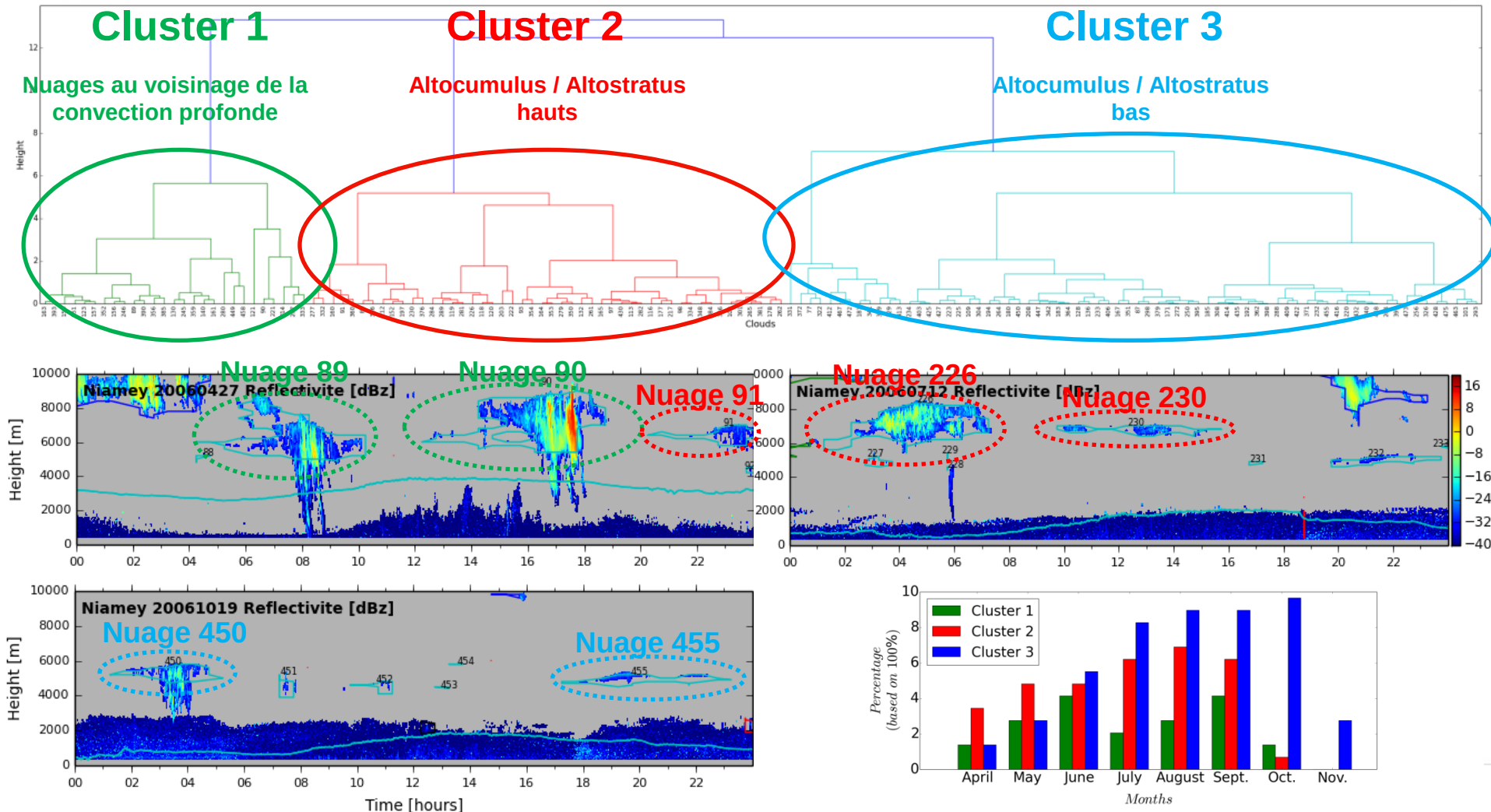
(pas JFM, durées > 30 min, détection radar $\geq 25\%$)



Utilisation des variables **bases** et **épaisseurs** de la base de données

Méthode de hiérarchisation Ward – Dendrogramme

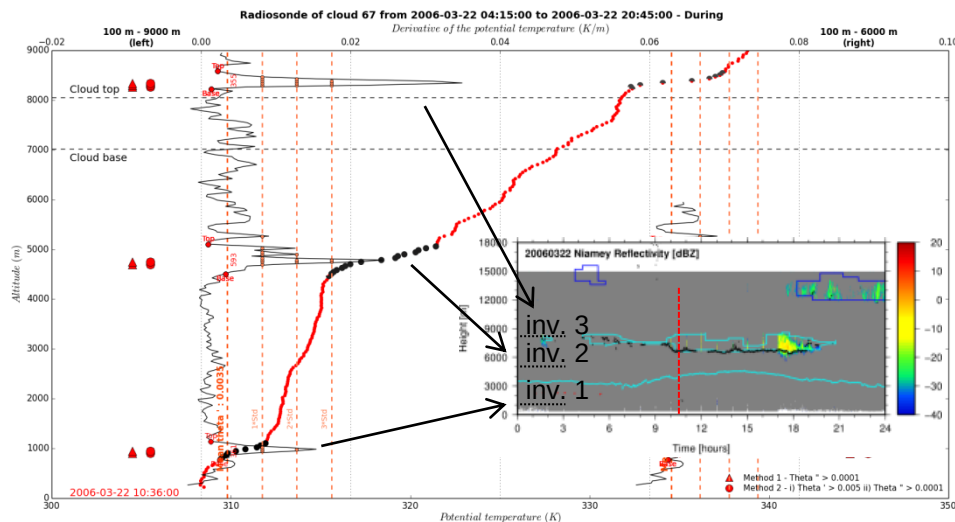
Station ARM Niamey (Niger) - 2006



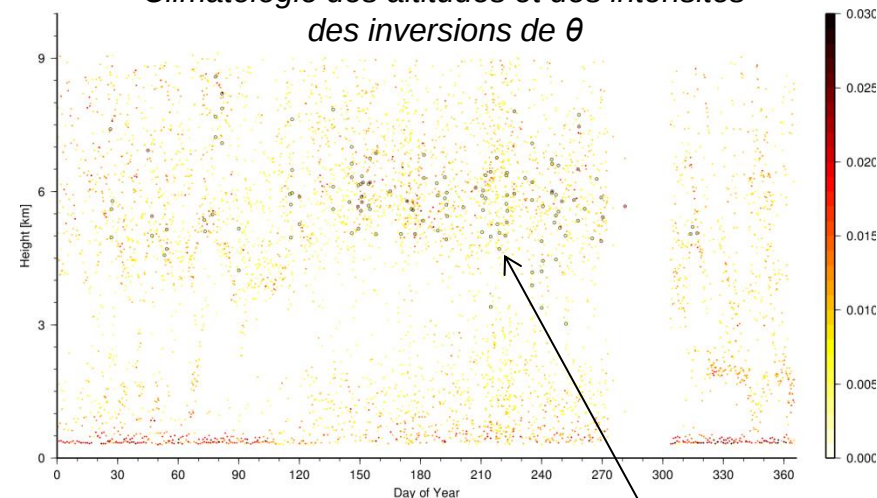
Cluster 1 : Pré et pendant la mousson, Cluster 2 : Pré et pendant la mousson,
Cluster 3 : Pendant et post mousson. Seul cluster avec des nuages en novembre.

Profils thermodynamiques – Détection des inversions de θ

Station ARM Niamey (Niger) - 2006



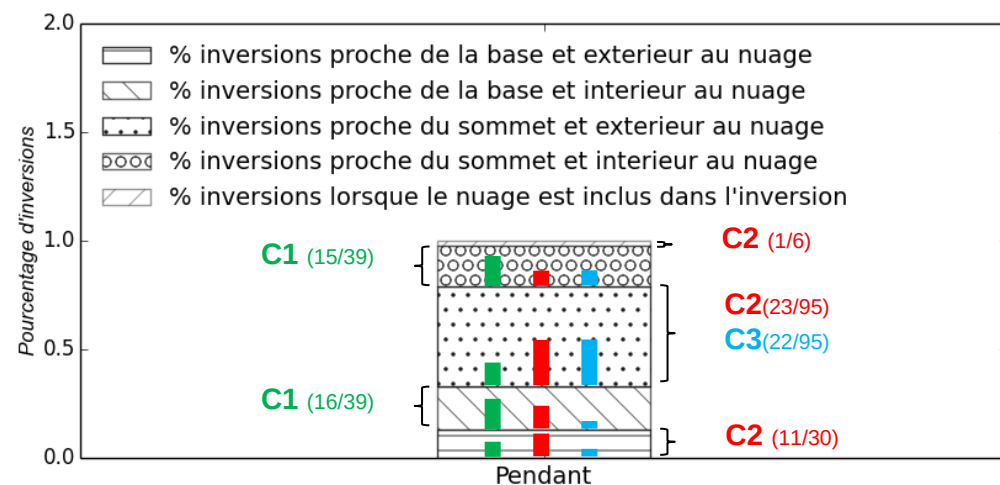
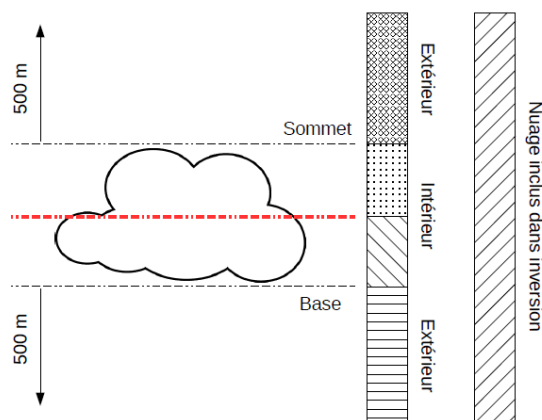
Climatologie des altitudes et des intensités des inversions de θ



Inversions avant formation du nuage de mi-niveau

322 nuages possèdent au moins un radiosondage

Positionnement des inversions de θ par rapport aux bases et aux sommets des nuages de mi-niveau



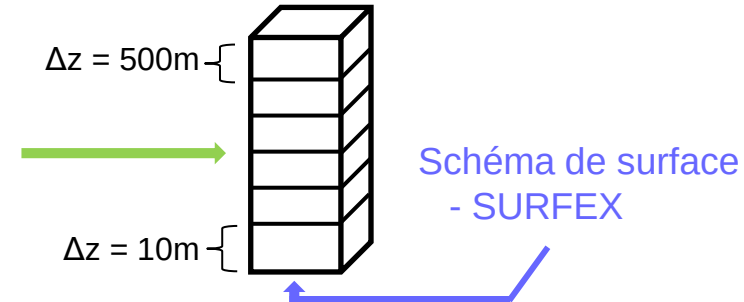
Simulation des nuages dans le modèle Més0-NH en 1D

Paramétrisations :

- Schéma turbulence et convection peu profonde : EDKF
- Schéma radiatif : RRTM/ECMWF
- Convection profonde : Kain-Fritsch
- Schéma microphysique : ICE3

Conditions initiales et forçages de grandes échelles :

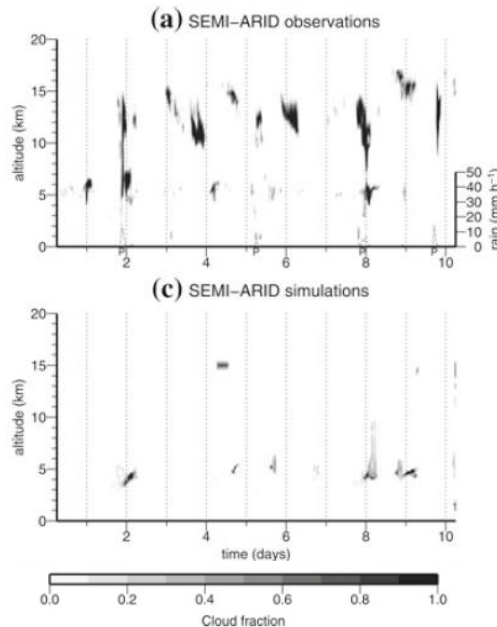
- Ré-analyse AMMA



Simulation 1D - ARM Niamey (Niger) - Régime semi-aride (Gounou et al., 2012)

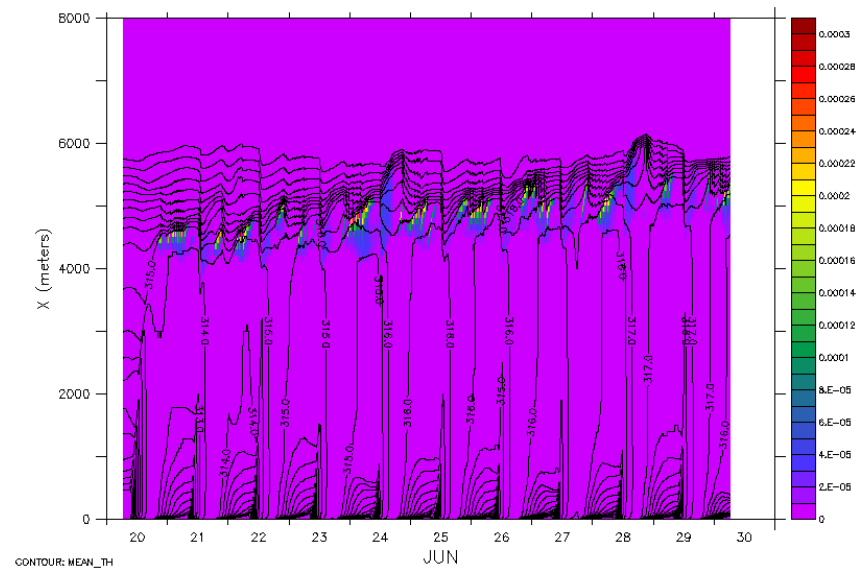
Du 20 au 30 juin 2006

10 * 1 jour



*Profils verticaux de fractions nuageuses simulées et observés
à la station ARM de Niamey sur 10*1 jour. (Couvreur et al., 2014)*

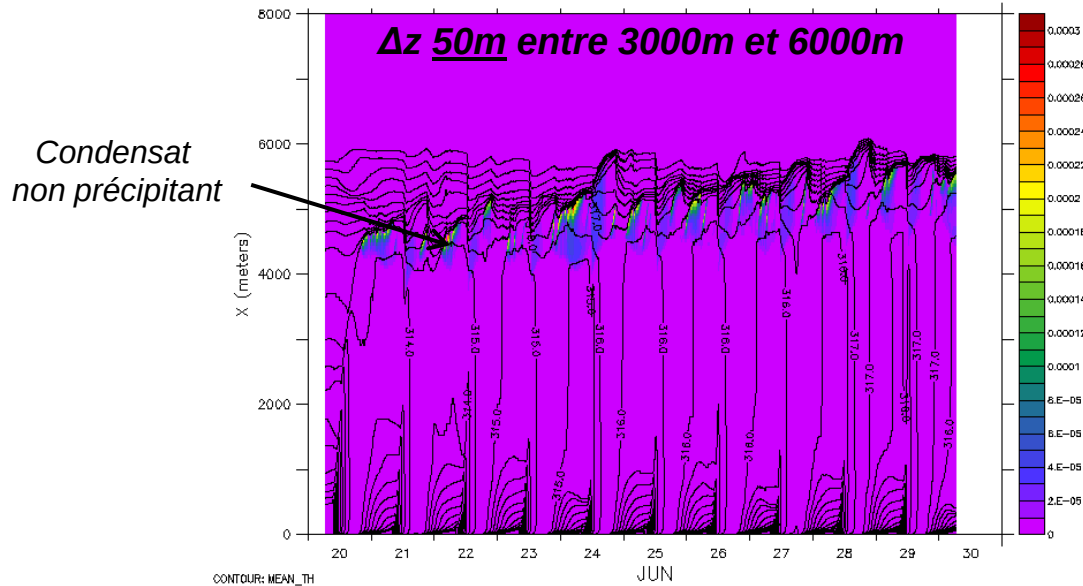
1 * 10 jours



*Profils verticaux de condensats non précipitants simulés sur 1*10 jours et profils de température potentielle associés.*

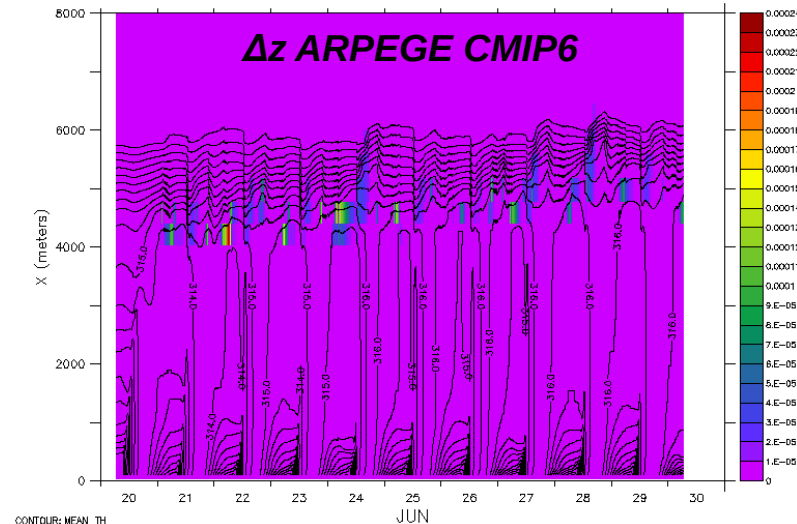
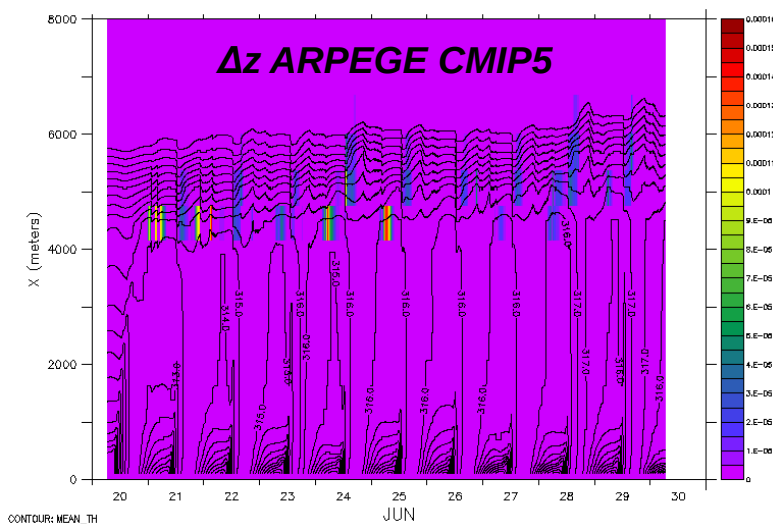
Simulation des nuages dans le modèle Més0-NH en 1D

Sensibilité à la résolution verticale (50m – 1000m)



Plus la résolution verticale est fine,

- plus la représentation des nuages est précise,
- plus l'altitude des sommets des nuages diminue et celle des bases augmente,
- plus la fraction d'eau liquide et de glace est importante aux sommets des nuages.



Représentation de la fraction d'eau liquide et de glace plus précise (résolution verticale plus fine).
Altitude des bases plus basse.

Conclusions et perspectives

Résumé et travaux en cours :

- ARM Niamey (Niger) 2006
- Analyse des **caractéristiques des nuages** de mi-niveau,
 - Identification des **différents types de nuages** de mi-niveau à partir des méthodes de clustering,
 - Analyse des **conditions thermodynamiques**,
 - Simulation des **nuages avec le modèle 1D Més0-NH** et de tests de sensibilité à la résolution verticale.

Perspectives :

- Caractéristiques
- ARM Niamey (Niger) 2006
- Pour chaque famille
- Réalisation des **composites de profils thermodynamiques** et de l'**impact radiatif** à la surface et au sommet de l'atmosphère,
 - **Documentation de l'environnement** (circulation atmosphérique, positionnement par rapport à l'ITD, etc.),
 - Etendre l'**étude** à l'**Afrique de l'Ouest** (données sols (FENNEC) et produits satellitaires (CloudSat/CALIPSO)).
- Processus
- Simulation des différents types de **nuages avec le modèle 1D Més0-NH** en se concentrant sur certains paramètres (profils thermodynamiques, cycle diurne, influence du vent, etc.).

Merci de votre attention

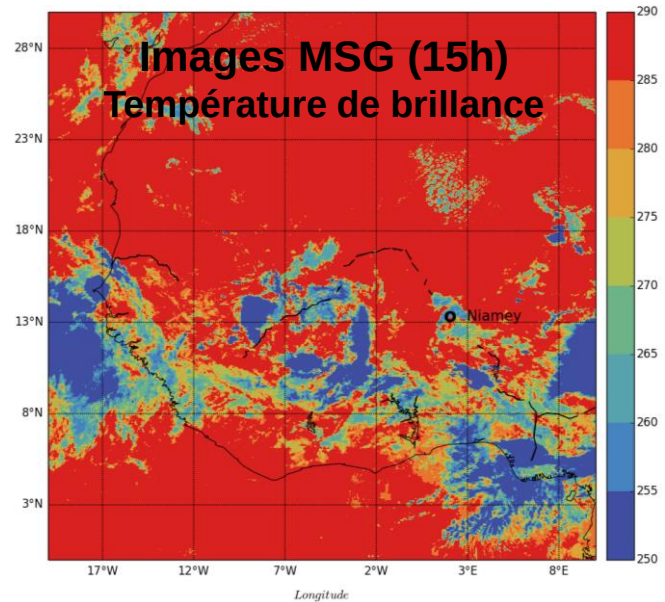
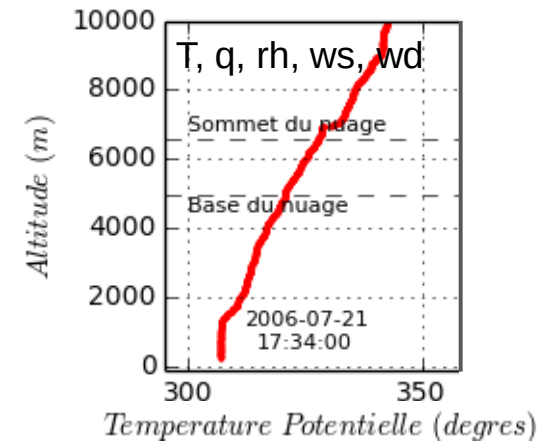
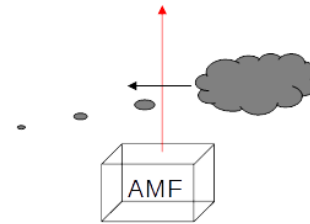
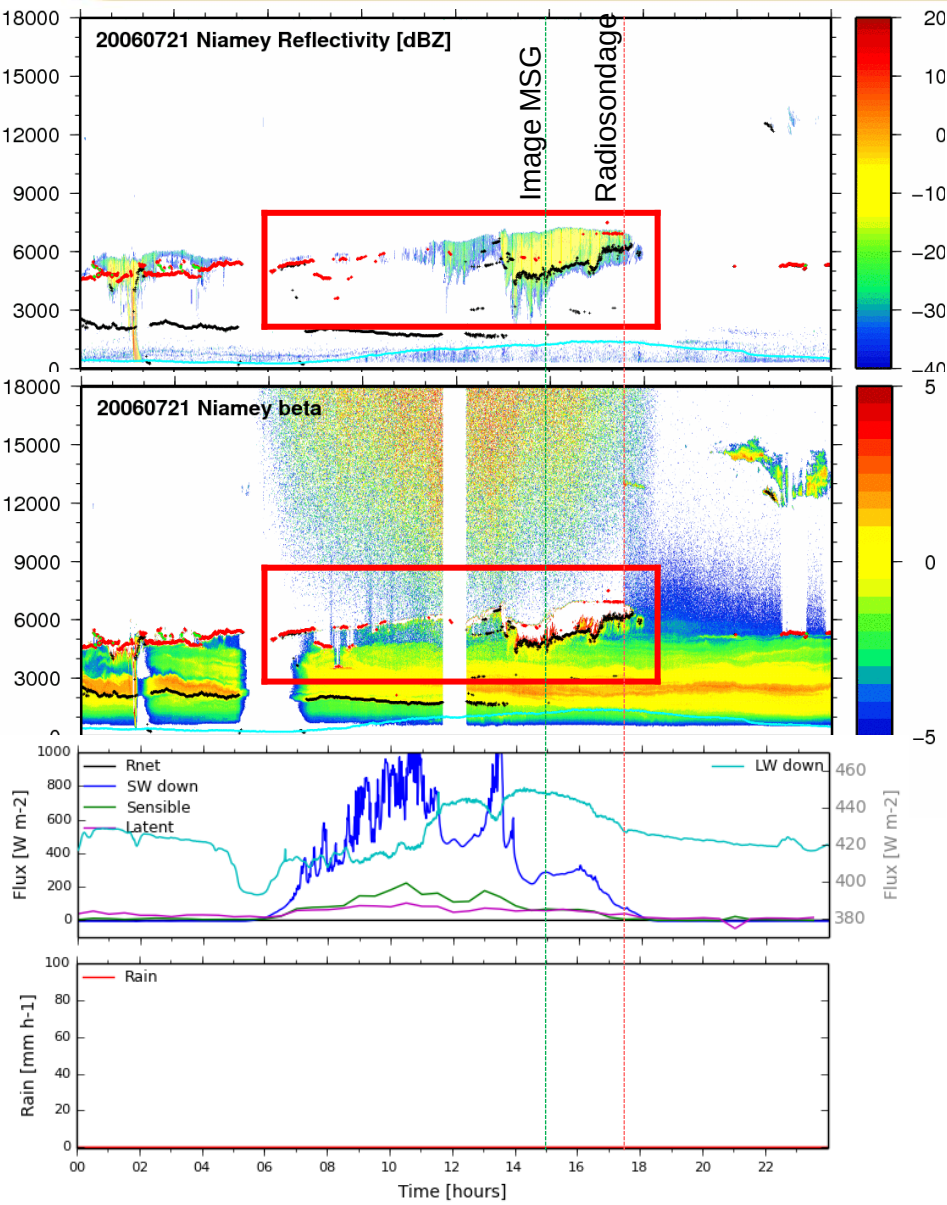
Annexes

Les jeux de données utilisés

Jeux de Données	Sols	Satellites
	AMMA (ARM Niamey)	Meteosat
		Meteosat
Année	2006	1981- -
Instruments	Lidar (L), radar (R), radiomètre (RM), pluviomètre (P), radiosondages (RS), pyrgéomètre (PG) et pyranomètre (PA)	Imageur
Profils verticaux de nébulosité (description couverture nuageuse)	X (R + L ou R ou L)	
Nuages		
- Masques nuageux (Occurrence)	X (R + L ou R ou L)	X
- Classification nuageuse	X (R + L + RM + P)	X
Fréquence temporelle (cycle diurne)	X (R + L ou R ou L)	X
Micro-physiques	X (R)	
Profils thermo-dynamiques	X (RS)	
Rayonnement		
- En surface	X (PG + PA)	X
- Au sommet		
Aérosols	X (L)	X

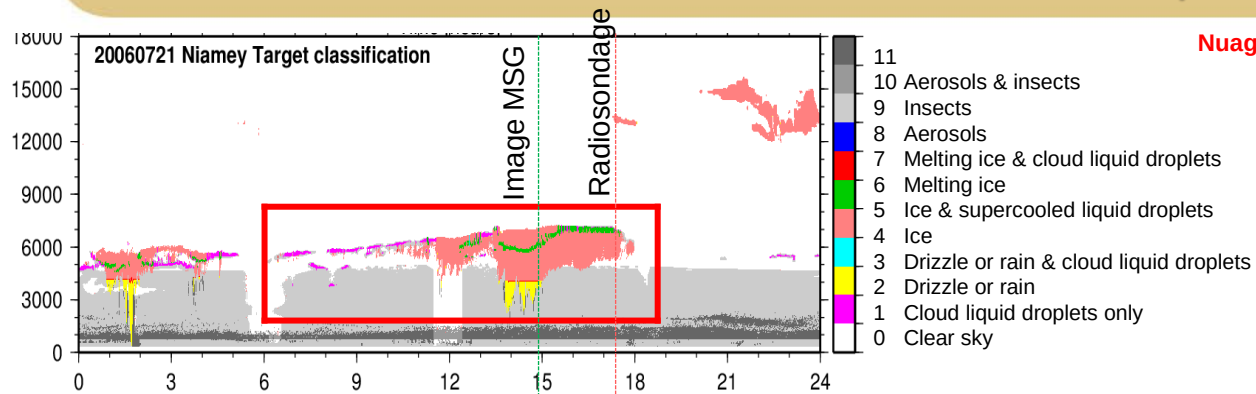
Méthodologie

Données radar et lidar brutes

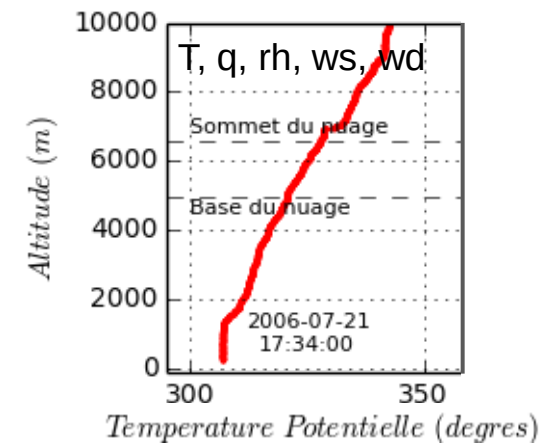


Méthodologie

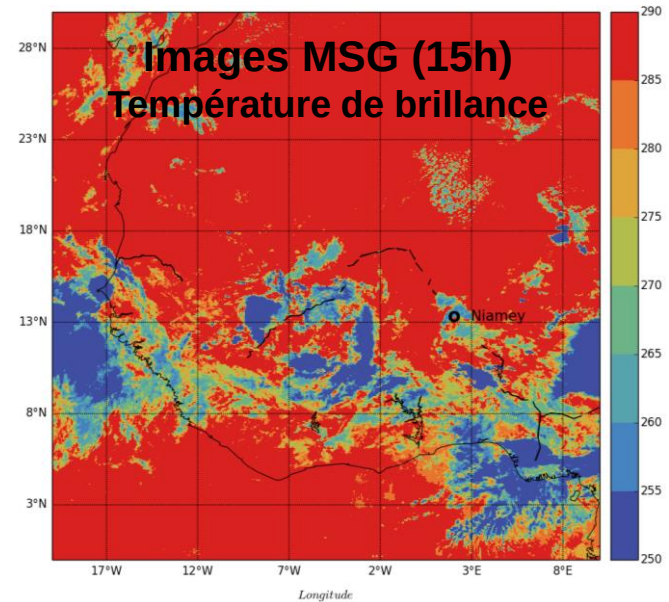
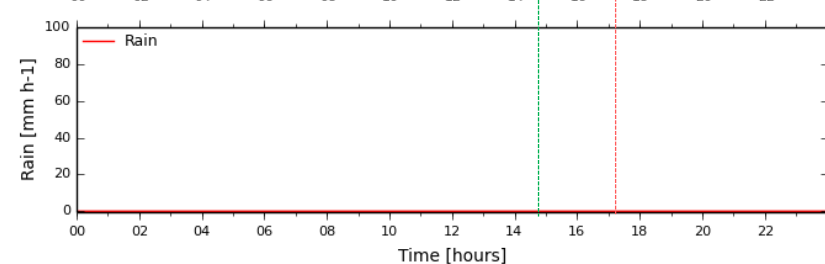
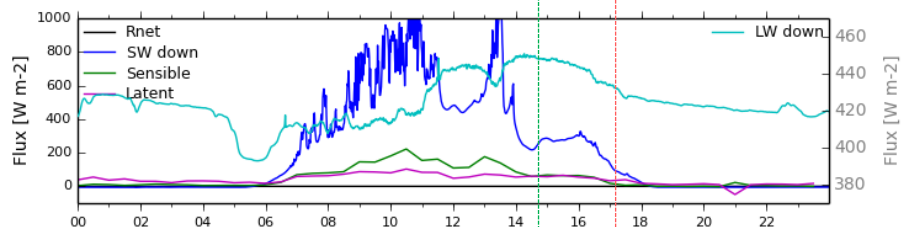
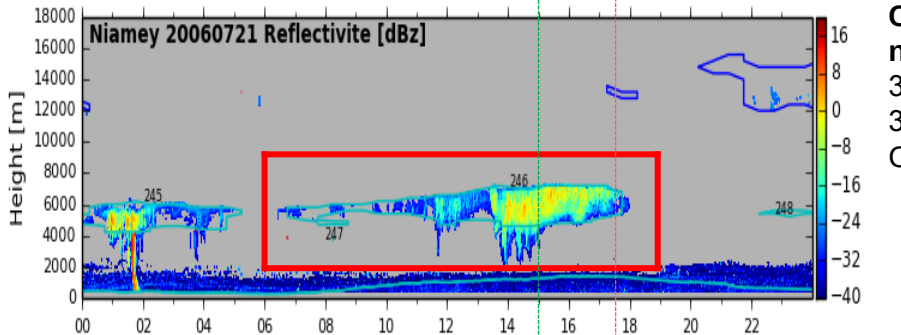
Classifications et masques



Nuages commencent à 1 (pour numérotation quicklooks)

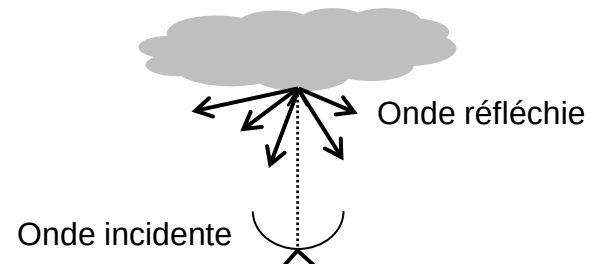


Critères définition nuages de mi-niveau :
 $3 \text{ km} \leq \text{altitude} \leq 7 \text{ km}$
 $3 \text{ km} \leq \text{base} \leq 7 \text{ km}$
 Composition : 1, 3, 4, 5, 6 ou 7

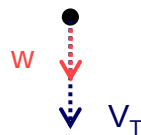


Les caractéristiques microphysiques

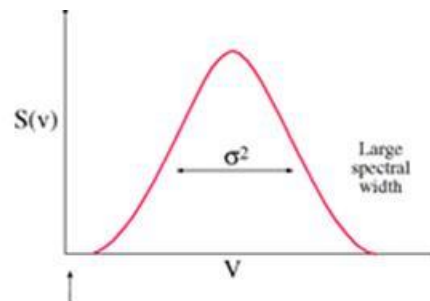
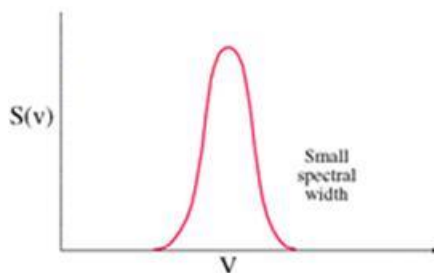
Réflectivité (dBz): rapport entre l'énergie électromagnétique réfléchiée et l'énergie électromagnétique incidente. Permet de fournir des informations sur la quantité d'eau sous forme liquide ou solide dans le volume sondé. Plus le contenu en eau est important, plus Z est grand.



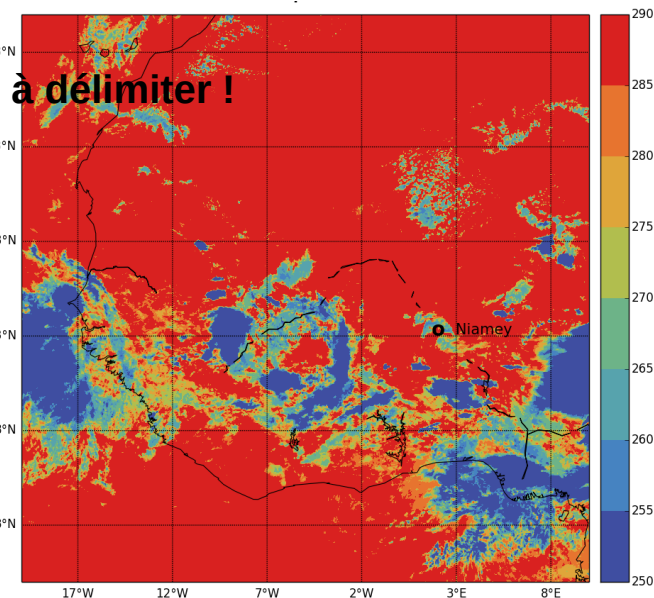
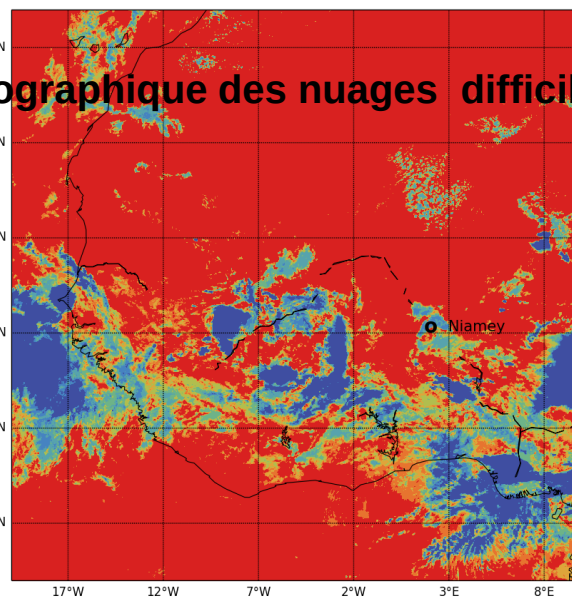
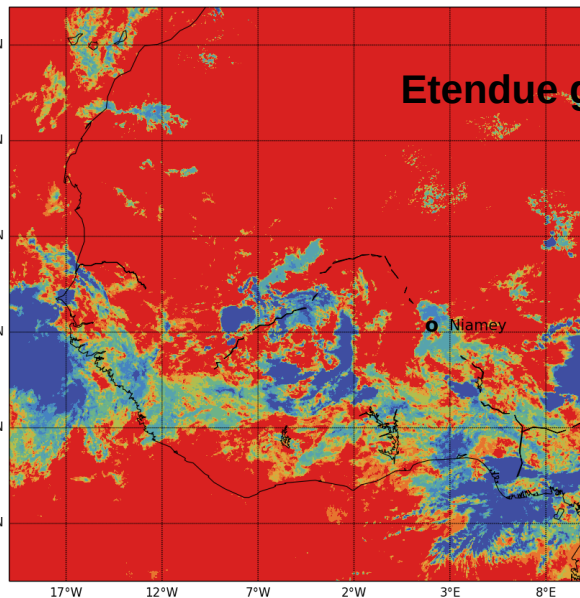
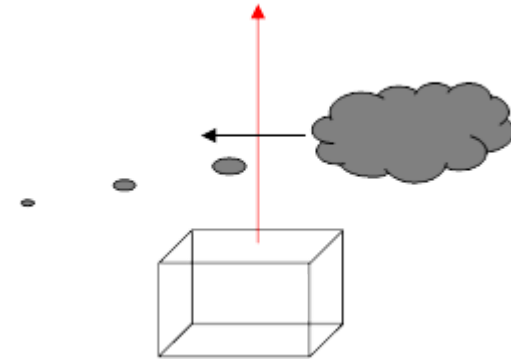
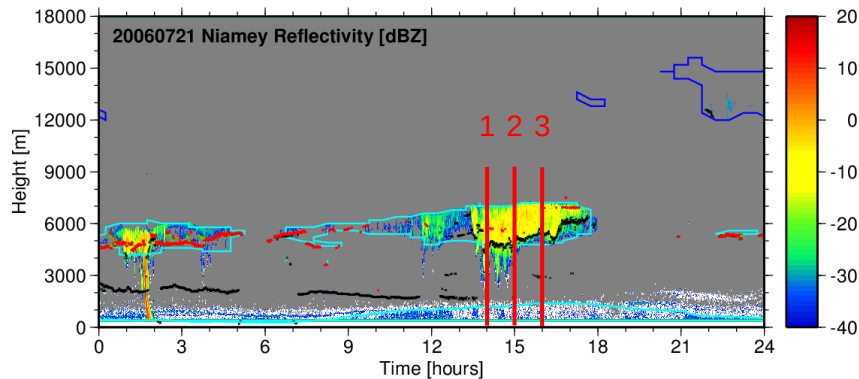
Vitesse Doppler (m.s^{-1}): elle permet de mesurer la vitesse de déplacement des cibles dans un volume donné. C'est la résultante de la vitesse verticale (w) et la vitesse terminale de chute des hydrométéores (= forme sous laquelle se présente l'eau) (V_T). Elle peut nous indiquer s'il y a ou pas des précipitations.



Spectral width (m.s^{-1}): c'est la variance du spectre Doppler. Elle s'obtient à partir de la variance due à la distribution dimensionnelle des hydrométéores.



Durée vs étendue géographique



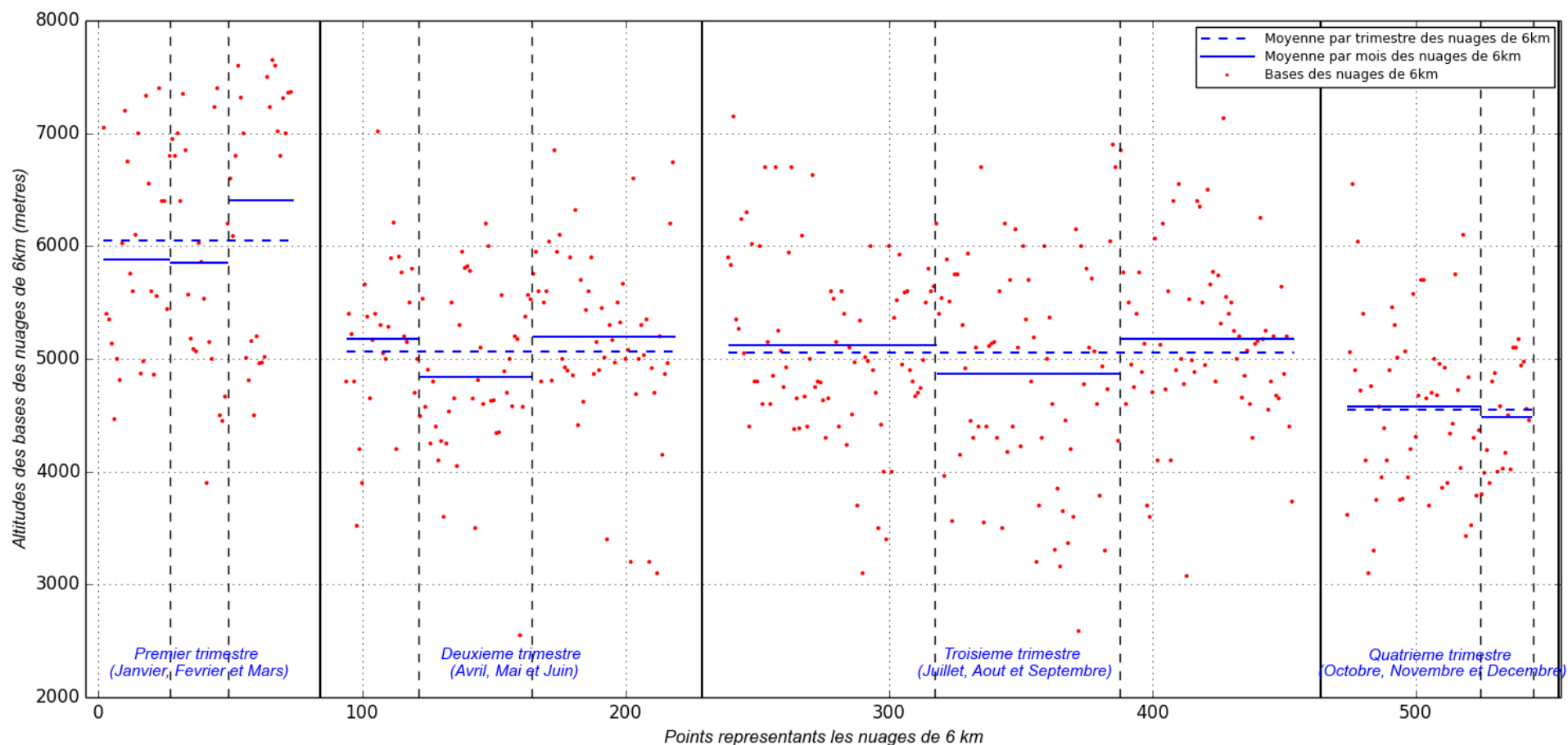
Etendue géographique des nuages difficile à délimiter !

Quicklooks du 21 juillet 2006 et vues satellitaires de la température de brillance (MSG)
au niveau de la région de Niamey du 21 juillet 2006 à 14h (1), 15h (2) et 16h (3)

Cycle saisonnier des bases des nuage

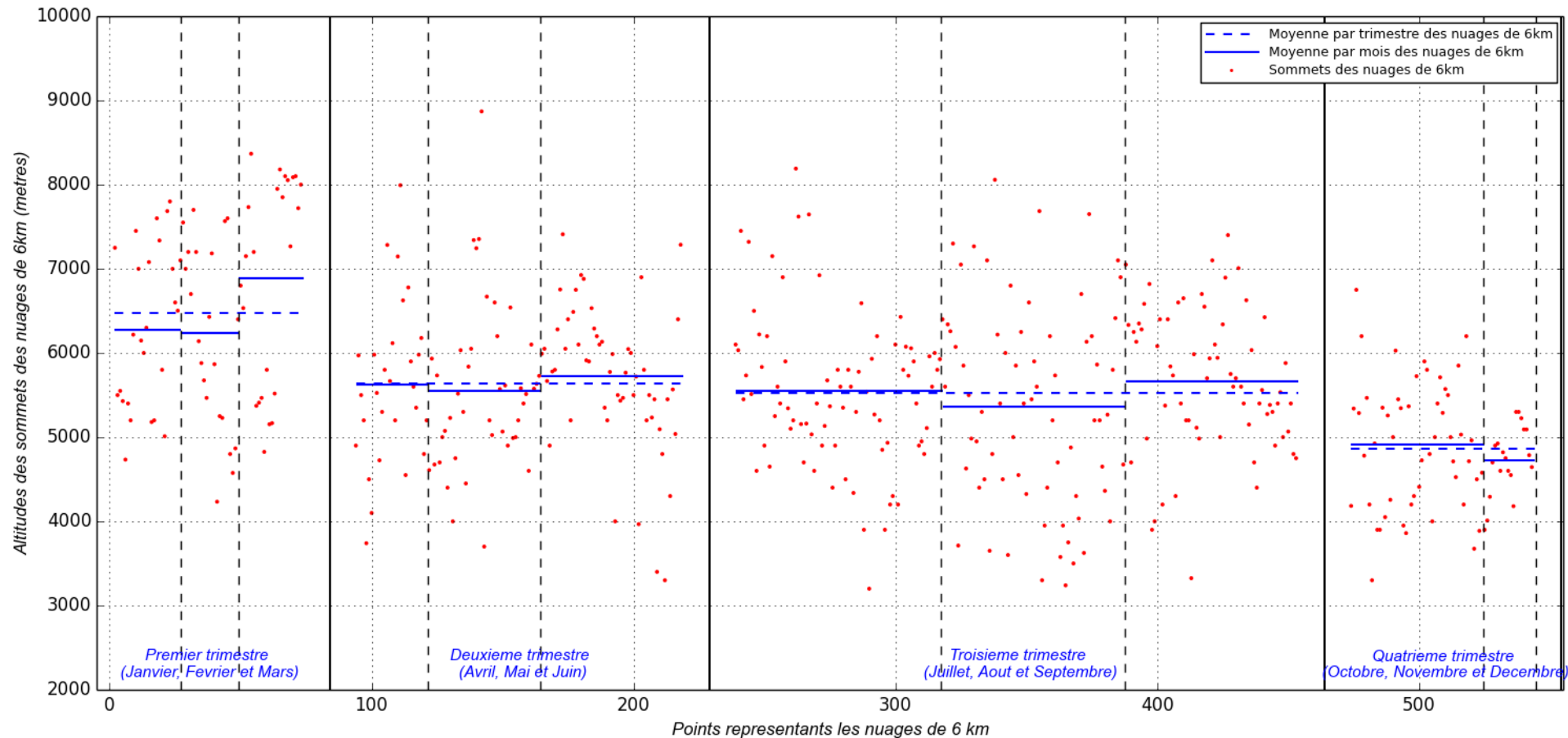
**482
nuages à
Niamey
en 2006**

Mois	J	F	M	A	M	J	J	A	S	O	N
Nombre de nuages	25	22	25	28	43	54	79	70	66	51	19



Cycle saisonnier des sommets

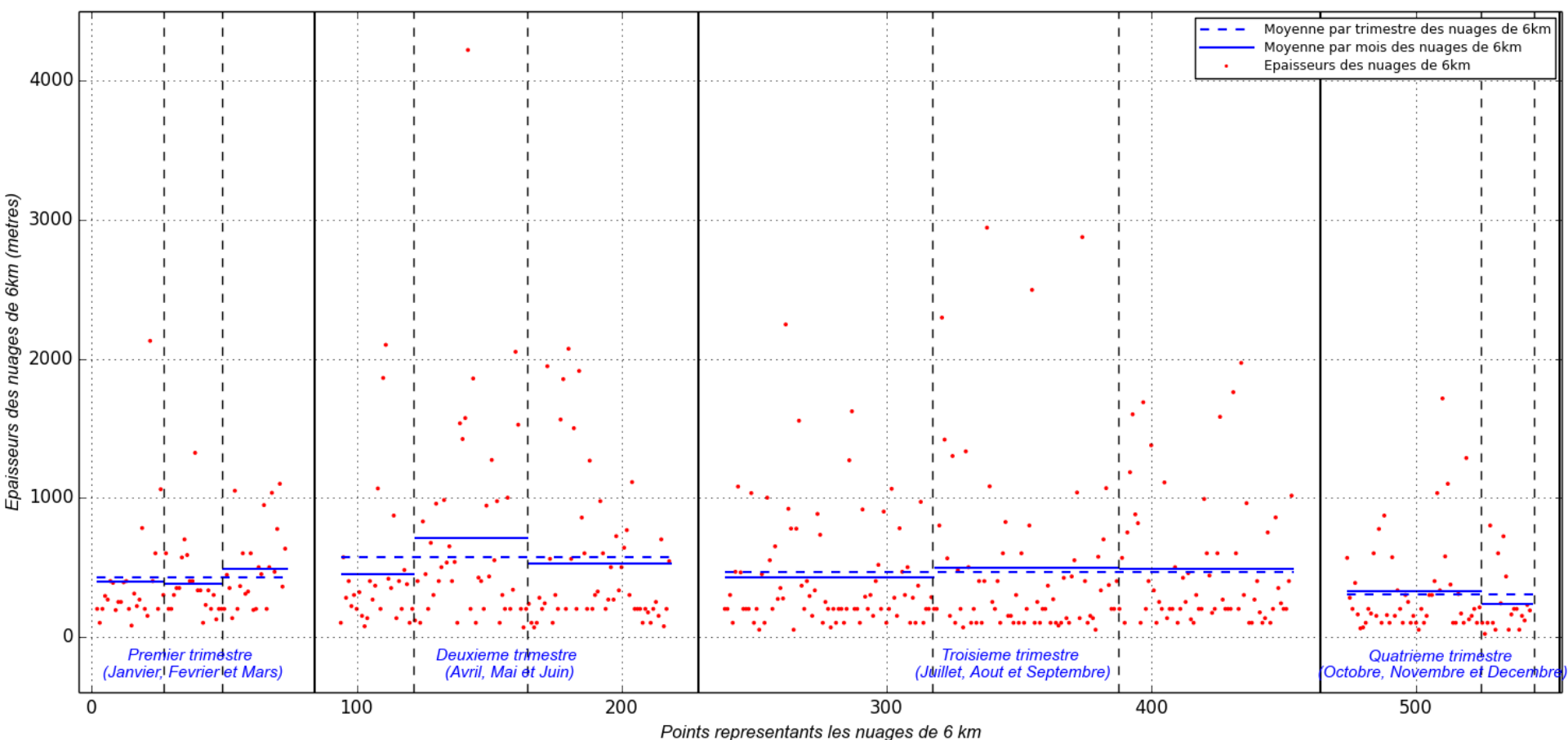
Mois	J	F	M	A	M	J	J	A	S	O	N
Nombre de nuages	25	22	25	28	43	54	79	70	66	51	19



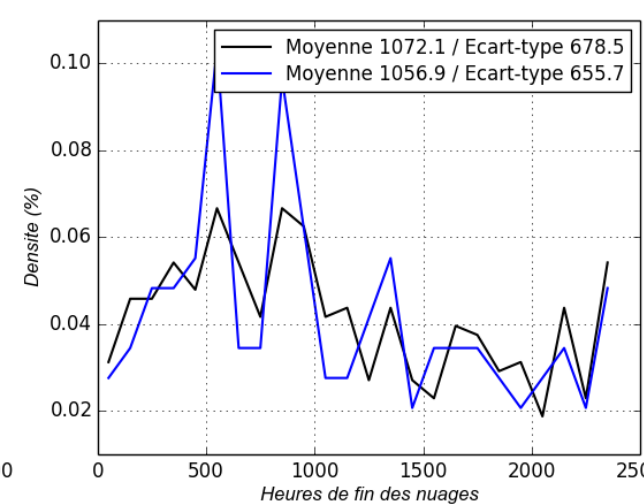
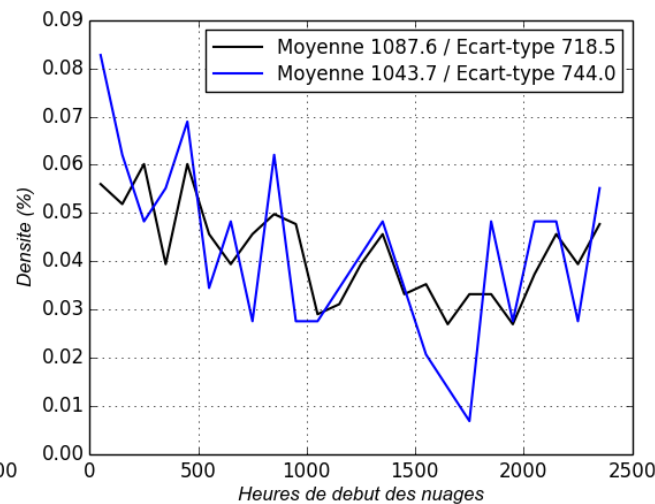
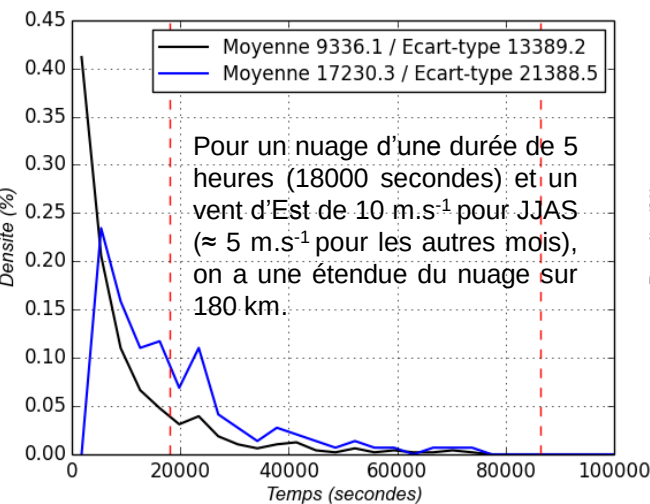
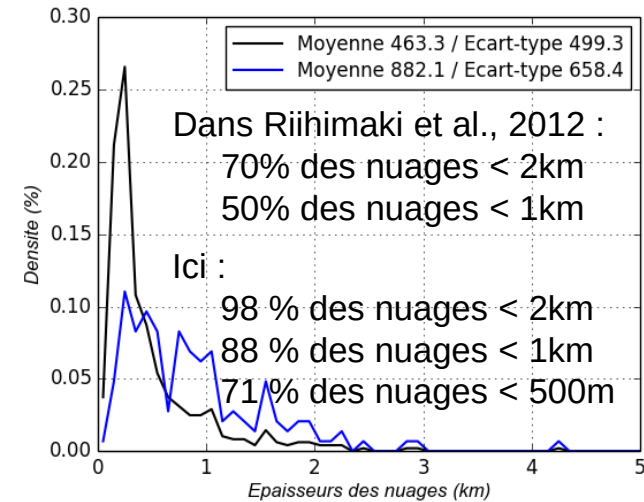
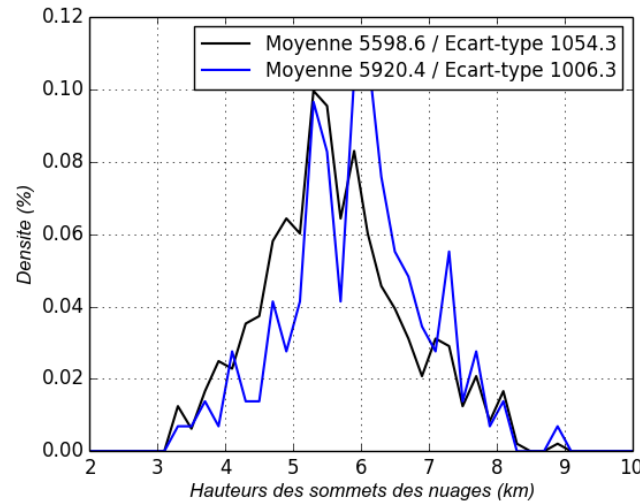
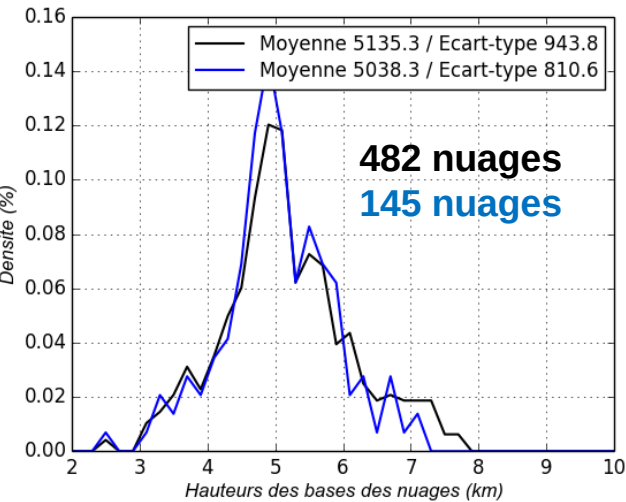
Cycle saisonnier des épaisseurs des nuages

**482
nuages à
Niamey
en 2006**

Mois	J	F	M	A	M	J	J	A	S	O	N
Nombre de nuages	25	22	25	28	43	54	79	70	66	51	19



Statistiques macrophysiques

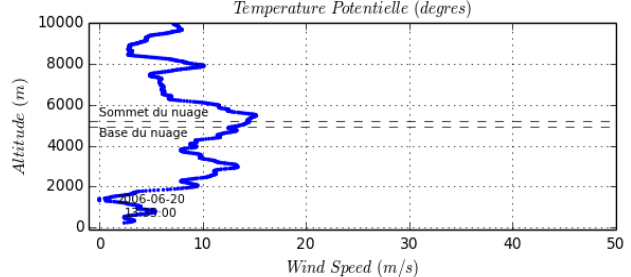
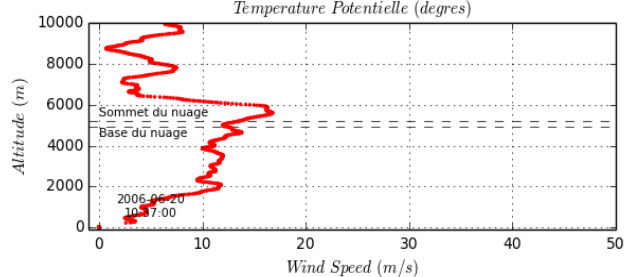
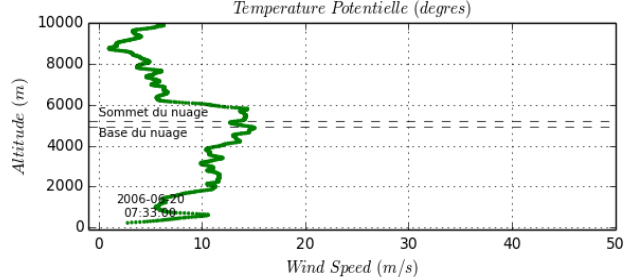
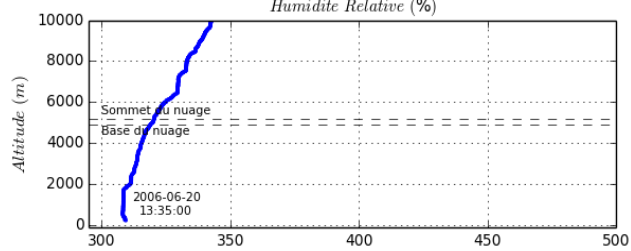
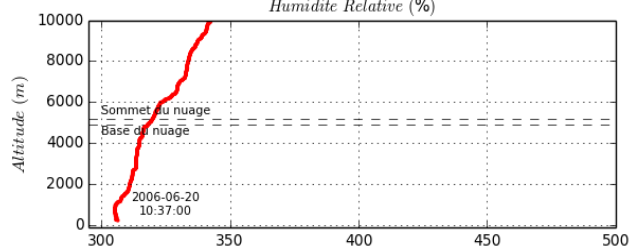
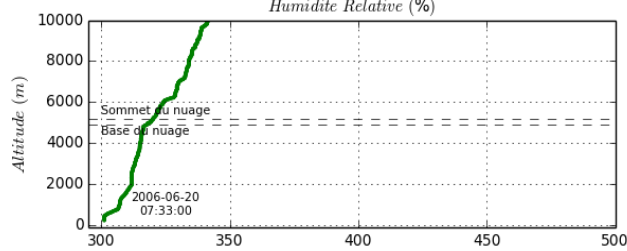
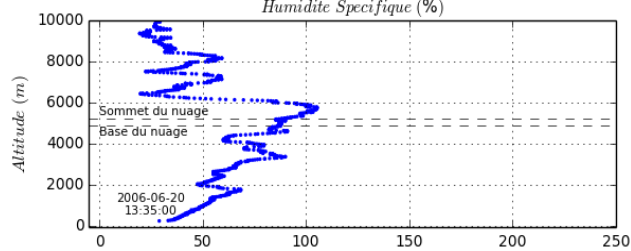
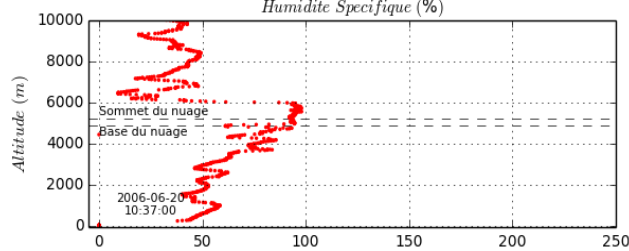
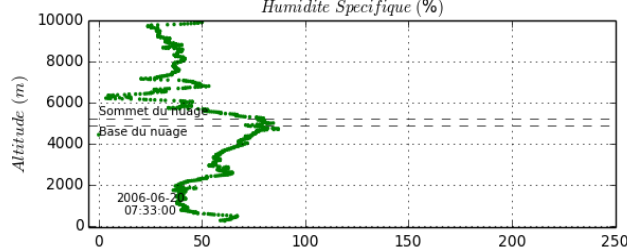
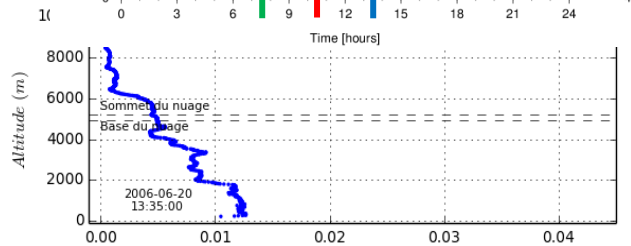
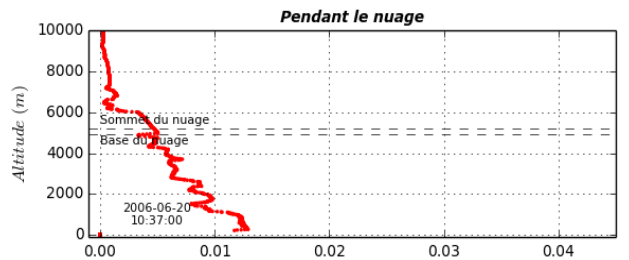
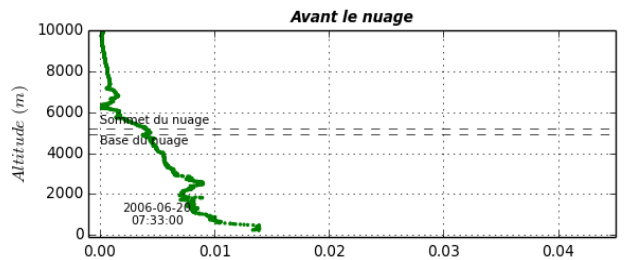
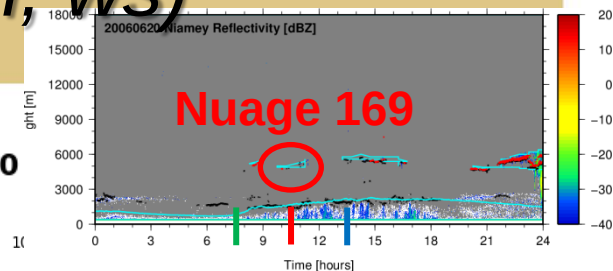


Histogrammes des caractéristiques des nuages de mi-niveau. En noir : tous les nuages de la base de données.
En bleu : pas JFM, durées > 30 min, détection radar >= 25%.

Radiosondages (T, qv, rh, ws)

Nuages commencent à 1

Radiosondages du nuage 169 du 2006-06-20 09:45:00 au 2006-06-20 11:15:00



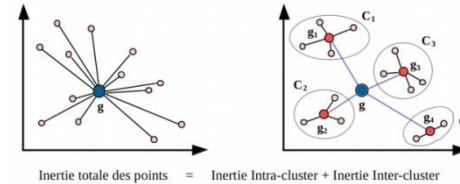
Les méthodes de clustering

Etude bibliographique - Qu'est-ce que c'est ? - Les différentes méthodes

Objectif : C'est une méthode d'analyse de données permettant de regrouper entre eux des objets similaires selon certains critères. Chaque groupe étant bien différencié des autres.

Finalité :

- Une grande similarité intra-groupe,
- Une faible similarité inter-groupe.



Besoin : La qualité d'un regroupement dépend de la mesure de similarité (= mesure de distance) utilisée par la méthode. Le choix de la mesure de distance entre individus dépend des objectifs et des données étudiées.

La distance Euclidienne
$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

Algorithmes principaux :

– Méthodes de partitionnement (non-déterministe):

Diviser les objets en un nombre fixe de groupes (K).

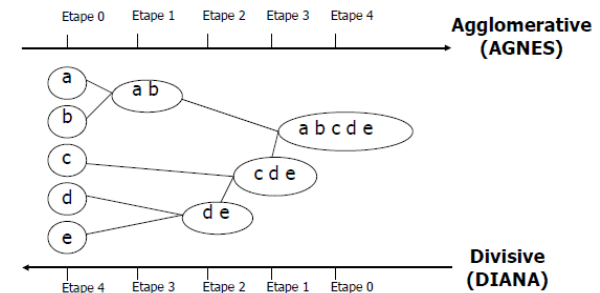
K-Means [k-moyennes] / K-Medoids [k-medoides] ou PAM [Partitioning Around Medoids] / Nuées dynamiques

– Méthodes hiérarchiques (déterministe) :

Produire une séquence de groupes imbriqués les uns dans les autres → dendrogramme.

Algorithmes agglomératifs [agglomerative]

Algorithmes de division [divisive]



Les méthodes de clustering

Etude bibliographique - Qu'est-ce que c'est ? - Les différentes méthodes

Objectif : C'est une méthode d'analyse de données permettant de regrouper entre eux des objets similaires selon certains critères. Chaque groupe étant bien différencié des autres.

Finalité :

- Une grande similarité intra-groupe,
- Une faible similarité inter-groupe.

Algorithmes principaux :

– Méthodes de partitionnement (non-déterministe):

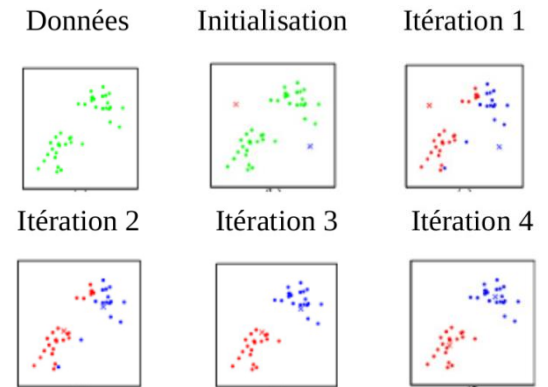
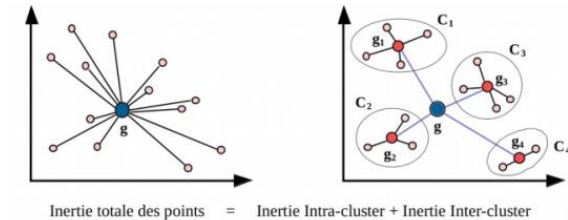
Diviser les objets en un nombre fixe de groupes (K).

K-Means [k-moyennes]

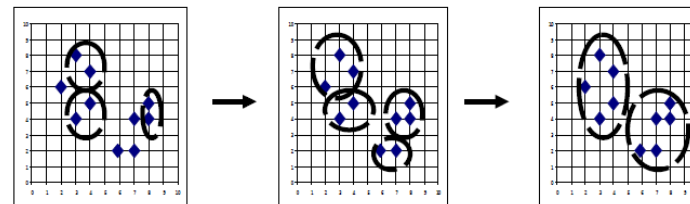
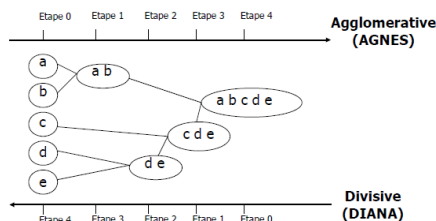
– Méthodes hiérarchiques (déterministe) :

Produire une séquence de groupes imbriqués les uns dans les autres → dendrogramme.

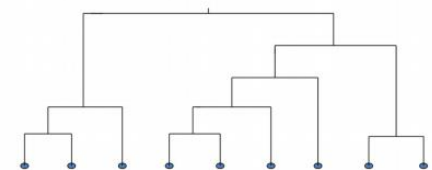
Algorithmes agglomératifs [agglomerative] - Méthode de Ward



Algorithme de K-Means



Algorithme agglomératif



Dendrogramme

Méthode de partitionnement K-Means (MacQueen, 1967)

1/ Caractéristiques générales de la méthode K-Means

- Chaque cluster est représenté par son centre,
- Chaque objet sera affecté au groupe dont le centre est le plus proche,
- Le centre d'un groupe est la moyenne de tous les points de ce groupe (barycentre),
- Méthode itérative (= jusqu'à ce que plus aucun élément ne bouge du cluster),
- Pas de cluster à contenu fixe,

Le barycentre de chaque groupe est recalculé à chaque nouvel individu introduit dans le groupe.

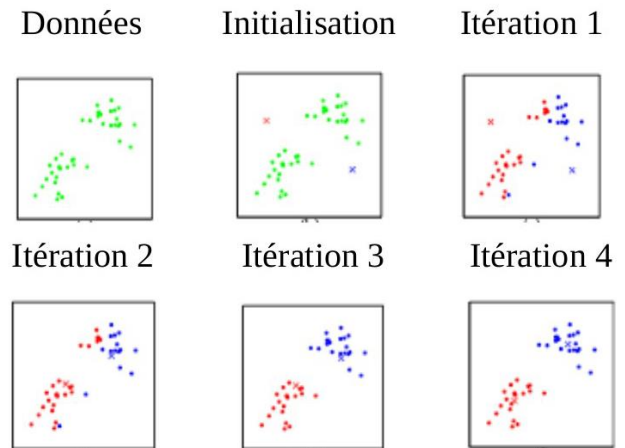
2/ La représentation schématique de K-Means avec $K = 2$ classes

3/ Avantages

- Simple,
- Rapide,
- Le plus utilisé tout en permettant de traiter de grandes bases de données.

4/ Inconvénients

- Spécification de K (nombre de clusters),
- Résultat dépend fortement du choix des centres initiaux, (ne fournit pas nécessairement un résultat optimum (i.e la partition en K groupes pour laquelle l'inertie intra-groupe est minimal mais un optimum local qui dépend des centres initiaux)),
- Mauvaise prise en compte des « outliers ».



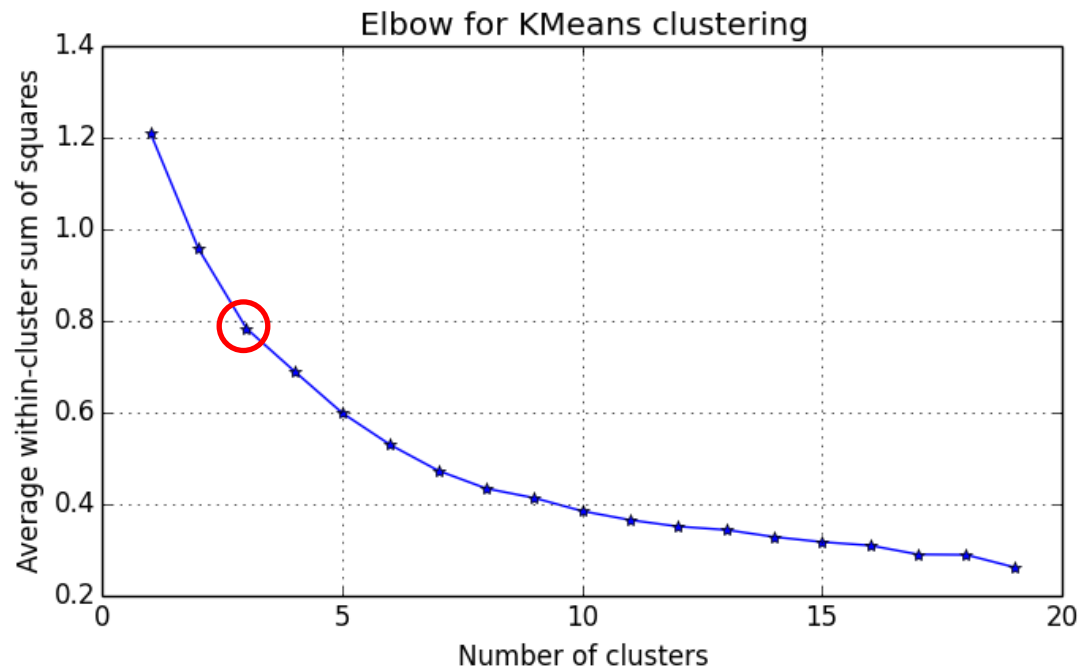
Les méthodes de clustering

Détermination du nombre de clusters K - Méthode Elbow

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar $\geq 25\%$))

La méthode Elbow nous renseigne sur la qualité des clusters pour un intervalle de possibilités de clusters. Un “bon” cluster veut dire que les éléments qu’il contient sont très rapprochés les uns des autres avec peu de dispersion.

Il faut choisir une valeur de K au point d’inflexion de la courbe.



Méthode de hiérarchisation AGNES - Ward

1/ Caractéristiques générales de la méthode hiérarchique AGNES

- Méthode ascendante,
- Regroupement par agglomération progressive d'éléments les plus proches deux à deux.

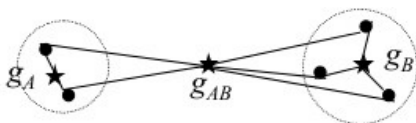
2/ Indice de dissimilarité

Distance Euclidienne

3/ Indice d'agrégation

Méthode de Ward (la plus connue et la plus courante)

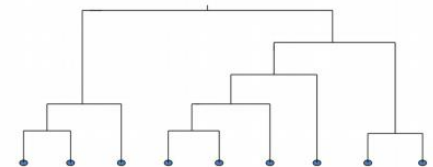
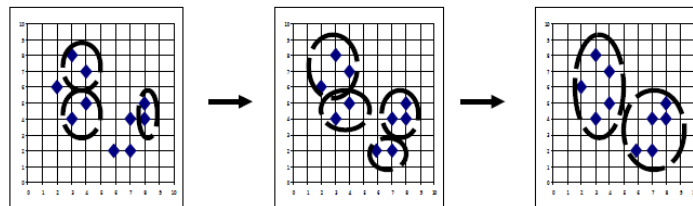
A chaque agrégation, gain minimum d'inertie intra-classe et perte d'inertie inter-classe



$$\delta(A, B) = \frac{p_A p_B}{p_A + p_B} d^2(g_A, g_B)$$

g_A = centre de gravité de la classe A (poids p_A)
 g_B = centre de gravité de la classe B (poids p_B)

3/ La représentation schématique de AGNES et du dendrogramme



4/ Avantages

- La plus utilisée,
- La lecture de l'arbre permet de déterminer le nombre optimal de classes,

5/ Inconvénients

- Lourdeur des calculs dès qu'on a un nombre de données important,
- Résultats différents en fonction de la paramétrisation (distances $\neq$, choix d'agrégation $\neq$)
- On peut trouver tout ce que l'on cherche $\rightarrow$ sur-interprétation des résultats.

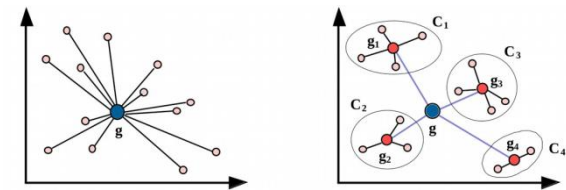
Les méthodes de clustering

Etude bibliographique - Qu'est-ce que c'est ?

C'est une méthode d'analyse de données permettant de regrouper entre eux des objets similaires selon certains critères. Les diverses techniques de classification visent toutes à répartir n individus, caractérisés par p variables $X_1, X_2, \dots, X_p$ en un certain nombre K de sous-groupes aussi homogènes que possible, chaque groupe étant bien différencié des autres.

Une bonne méthode de regroupement permet de garantir,

- Une grande similarité intra-groupe,
- Une faible similarité inter-groupe.



Bonne partition (Bisson, 2001)

Le clustering (= regroupement) est de la classification non supervisée, il n'y a pas de classes prédéfinies.

La qualité d'un regroupement dépend donc de la mesure de similarité utilisée par la méthode. La similarité est exprimée par le biais d'une mesure de distance. Les définitions de distance sont très différentes que les variables soient des données continues sur un intervalle (distances de Minkowski, de Manhattan et Euclidienne), nominales ou booléennes. Le choix de la mesure de distance entre individus dépend donc des objectifs et des données étudiées.

La distance Euclidienne
$$d(i, j) = \sqrt{(|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + \dots + |x_{ip} - x_{jp}|^2)}$$

Matrice de données

$$\begin{bmatrix} x_{11} & \dots & x_{1f} & \dots & x_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{if} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nf} & \dots & x_{np} \end{bmatrix}$$

Matrice de distances
(= mesure de dissimilarité)

$$\begin{bmatrix} 0 & & & & \\ d(2,1) & 0 & & & \\ d(3,1) & d(3,2) & 0 & & \\ \vdots & \vdots & \vdots & & \\ d(n,1) & d(n,2) & \dots & \dots & 0 \end{bmatrix}$$

Les méthodes de clustering

Etude bibliographique - Les différentes méthodes de clustering

Il y a **deux classes principales d'algorithmes** :

– Méthodes de partitionnement :

Diviser de manière optimale l'ensemble des objets en un nombre fixe de groupes (K).
L'objectif est ici de minimiser une mesure de la dissemblance intra-groupe pour K groupes.
Le problème étant lié à l'optimisation d'une combinatoire, la solution trouvée sera rarement l'optimum global mais plutôt un des nombreux optimums locaux.

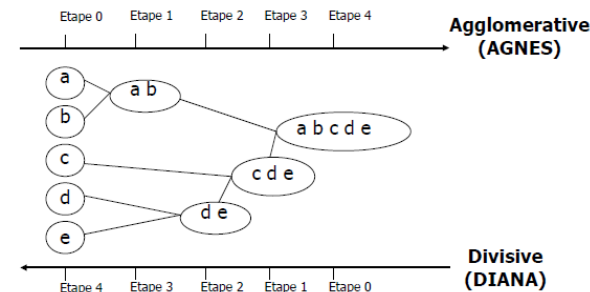
K-Means [k-moyennes] / K-Medoids [k-medoides] ou PAM [Partitioning Around Medoids] / Nuées dynamiques

– Méthodes hiérarchiques :

Produire une séquence de groupes imbriqués les uns dans les autres.
La plupart des méthodes de groupement hiérarchique produisent une représentation des données de type dendrogramme.

Algorithmes agglomératifs [agglomerative]

Algorithmes de division [divisive]



– Méthodes basés sur la densité :

Grouper les objets tant que la densité de voisinage excède une certaine limite.

DBSCAN

– Méthodes basés sur des grilles :

Diviser l'espace en cellules formant une grille multi-niveaux et grouper les cellules voisines en terme de distance.

Les méthodes de clustering

Etude bibliographique - La méthode de partitionnement K-Means (MacQueen, 1967)

Objectif : Construire une partition à K clusters d'une base D de n objets.

1/ Caractéristiques générales de la méthode K-Means

- Chaque cluster est représenté par son centre,
- Chaque objet sera affecté au groupe dont le centre est le plus proche,
- Le centre d'un groupe est la moyenne de tous les points de ce groupe (barycentre),
- Pas de cluster à contenu fixe.

2/ L'algorithme K-Means

1. Fixer le nombre de cluster K ,
2. Choisir aléatoirement K centres pour chaque cluster,
3. Affecter chaque objet O au cluster C_i de centre M_i tel que $\text{dist}(O, M_i)$ soit minimale,
4. Calculer M_i de chaque cluster (ie. le barycentre),
5. Refaire les étapes 3 et 4 jusqu'à ce qu'il y ait convergence, c'est-à-dire jusqu'à ce qu'il n'y ait plus aucun individu à changer de classe ; donc que les groupes ne bougent plus.

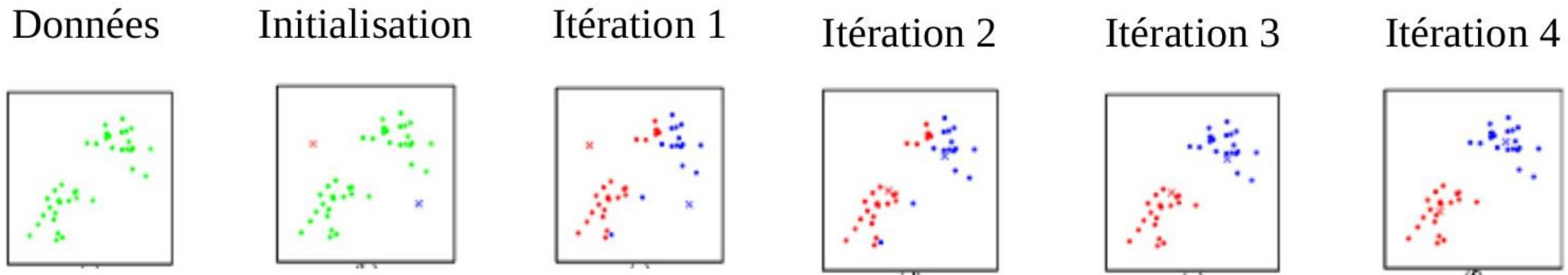
Le barycentre de chaque groupe est recalculé à chaque nouvel individu introduit dans le groupe, au lieu d'attendre l'affectation de tous les individus.

C'est une **méthode itérative**, jusqu'à ce que plus aucun élément ne bouge du cluster. L'inertie intra-classe diminue à chaque étape.

Les méthodes de clustering

Etude bibliographique - La méthode de partitionnement K-Means (MacQueen, 1967)

3/ La représentation schématique de K-Means avec $K = 2$ classes



4/ Avantages

- Simple et rapide (convergence parfois possible en une seule itération donc relativement efficace),
- Le plus utilisé tout en permettant de traiter de grandes bases de données.

5/ Inconvénients

- Spécification de K (nombre de clusters),
- Le résultat dépend fortement du choix des centres initiaux, (ne fournit pas nécessairement un résultat optimum (i.e la partition en K groupes pour laquelle l'inertie intra-groupe est minimal mais un optimum local qui dépend des centres initiaux)),
- Algorithme **non-déterministe** : il produit un résultat différent à chaque exécution car les centres au départ de l'algorithme sont choisis aléatoirement,
- Mauvaise prise en compte des "outliers" points extrêmes en dehors des groupes qui faussent les moyennes et donc les centres.

Les méthodes de clustering

Etude bibliographique - La méthode de hiérarchisation AGNES (Ward)

Objectif : Créer une décomposition hiérarchique en groupes (clusters). Utiliser la matrice de distances comme critère de regroupement. K n'a pas à être précisé, mais a besoin d'une condition d'arrêt.

1/ Caractéristiques générales de la méthode hiérarchique AGNES

- Méthode ascendante,
- Regroupement par agglomération progressive d'éléments les plus proches deux à deux,
- Fusionne les nœuds qui ont la plus faible dissimilarité,
- On peut se retrouver dans la situation où tous les nœuds sont dans le même groupe,
- Nécessiter de choisir un indice de dissimilarité entre les individus et un indice d'agrégation.

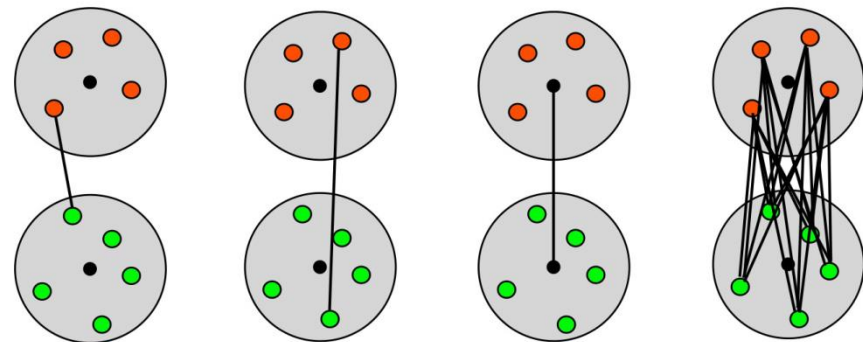
2/ Indice de dissimilarité

- Distance de Minkowski,
- Distance de Manhattan,
- Distance Euclidienne,
- Etc.

3/ Indice d'agrégation

- Différents critères
donc différentes hiérarchies

Distance inter-groupe : type de lien [linkage]



Groupe à liens simples
[single/closest neighbors]

Groupe à liens complets
[complete/furthest neighbors]

Groupe à centroïde
(centroid) †

Groupe à liaison moyenne
(average/mean/UPGMA) †

Les méthodes de clustering

Etude bibliographique - La méthode de hiérarchisation AGNES (Ward)

3/ Indice d'agrégation (suite)

Méthode de Ward (la plus connue et la plus courante)

Objectif : Gain minimum d'inertie intra-classe à chaque agrégation et perte d'inertie inter-classe due à cette agrégation.

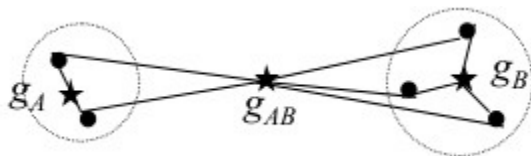
Elle consiste à réunir les deux clusters dont le regroupement fera le moins baisser l'inertie inter-classe. La distance de Ward entre deux classes est celle de leurs barycentre au carré, pondérée par les effectifs des deux clusters. Cette technique tend à regrouper les petites classes entre elles.

$$\delta(A, B) = \frac{p_A p_B}{p_A + p_B} d^2(g_A, g_B)$$

g_A = centre de gravité de la classe A (poids p_A)

g_B = centre de gravité de la classe B (poids p_B)

Elle constitue une des meilleures méthodes de classification hiérarchique sur données euclidiennes.



$$I_{AB} - (I_A + I_B)$$

inertie

Les méthodes de clustering

Etude bibliographique - La méthode de hiérarchisation AGNES (Ward)

3/ L'algorithme de AGNES

Partir d'un seul objet auquel on associe les autres objets un par un.

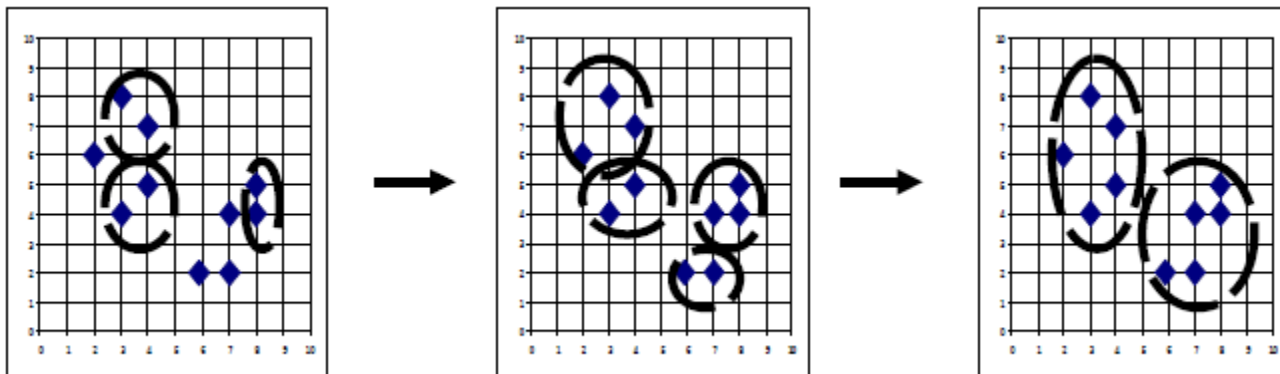
Initialisation

1. Chaque individu est placé dans son propre cluster,
2. Calcul de la matrice de ressemblance (= matrice des distances) M entre chaque couple de clusters (entre chaque individu deux a deux).

Répéter

3. Sélection dans M des deux clusters les plus proches C I et C J. Ils forment un « sommet ».
4. Fusion de C I et C J pour former un cluster C G,
5. Mise à jour de M en calculant à l'aide d'un critère d'agrégation la ressemblance entre C G et les clusters existants.

Jusqu'à la fusion des 2 derniers clusters en un seul.



Les méthodes de clustering

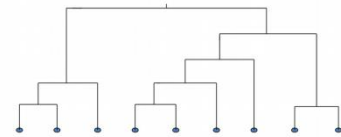
Etude bibliographique - La méthode de hiérarchisation AGNES (Ward)

6/ Le dendrogramme ou l'arbre hiérarchique

- Sur le dendrogramme, la position de deux objets au sein d'un groupe ou entre les groupes n'est pas liée à leur proximité dans l'espace des données,
- Ce qui importe ce sont les relations hiérarchiques entre les groupes,
- La longueur/hauteur d'une branche est proportionnelle à la distance entre deux classes,
- En traçant une ligne verticale/horizontale à une hauteur donnée de l'arbre, on peut identifier K groupes.

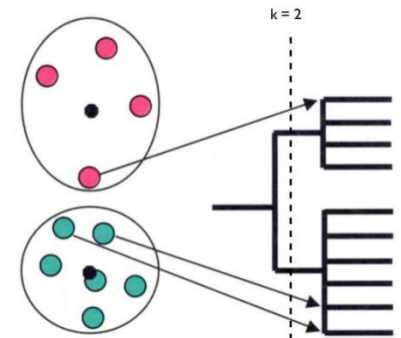
Coupure de l'arbre, pour déterminer une partition intéressante, à opérer :

- * après la série d'agrégation à faibles distances (branches courtes de l'arbre),
- * avant les agrégations à grandes distances (branches longues).



7/ Avantages

- Méthode hiérarchique la plus utilisée,
- La lecture de l'arbre permet de déterminer le nombre optimal de classes,
- Algorithme **déterministe**.

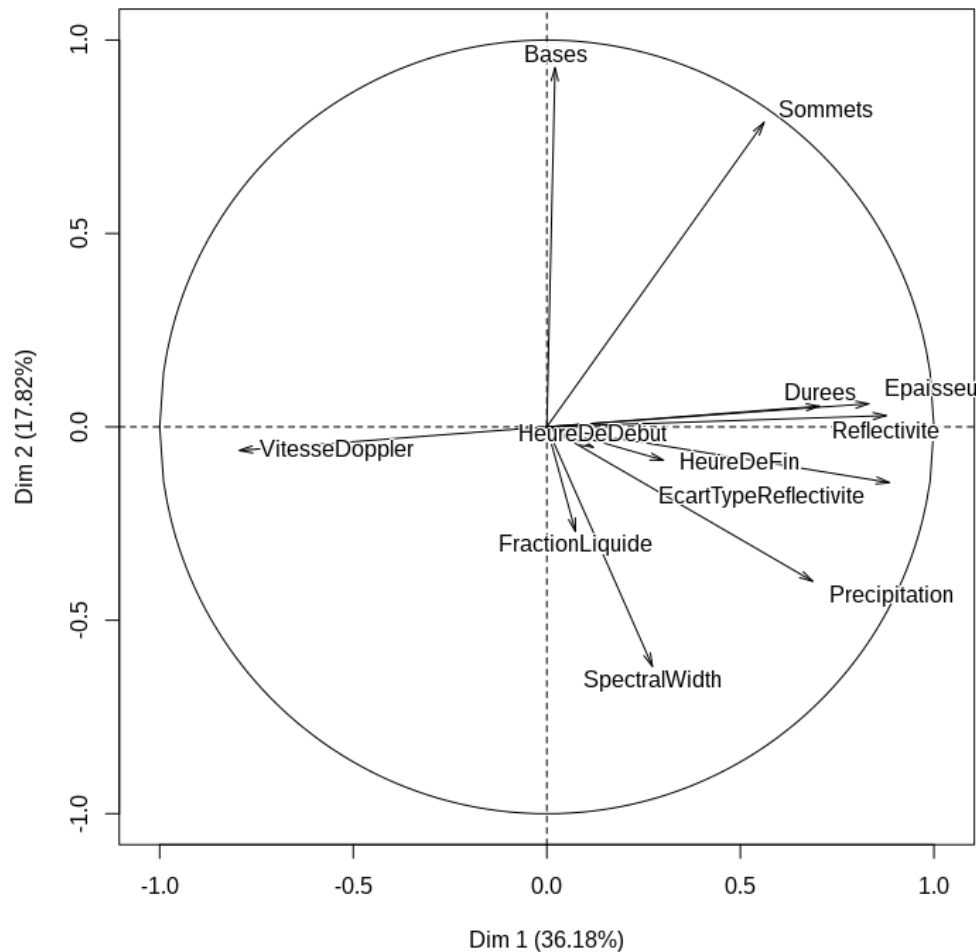


8/ Inconvénients :

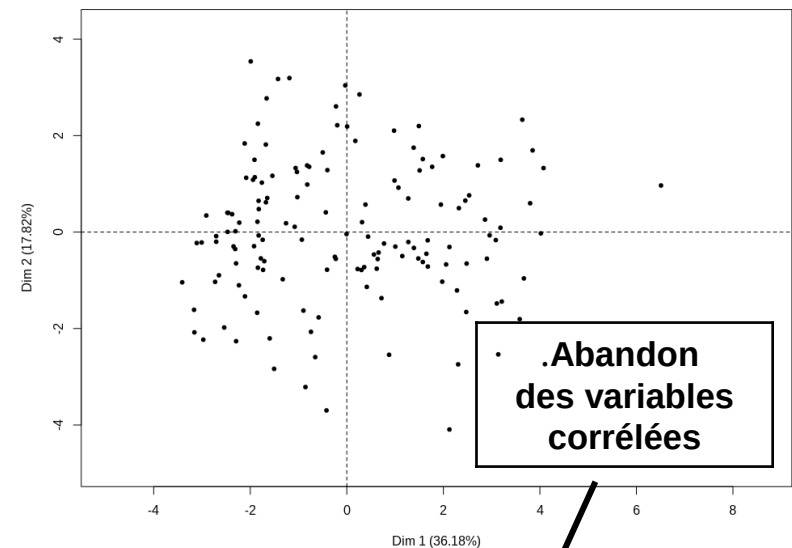
- Plus on est haut dans l'arbre, moins bonne est la représentation de la structure des données,
- Lourdeur des calculs dès qu'on a un nombre de données important
→ coûteux en temps de calcul,
- On peut trouver tout ce que l'on cherche (cf. ≠ critères d'agrégation)
→ attention à la sur-interprétation des résultats,
- Résultats différents en fonction de la paramétrisation (distances ≠, choix d'agrégation ≠)

Analyses en Composantes Principales (ACP)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Dim 1	Dim 2	Dim 3	Dim 4	Total
36,18%	17,82%	11,44%	10,23%	75,67%

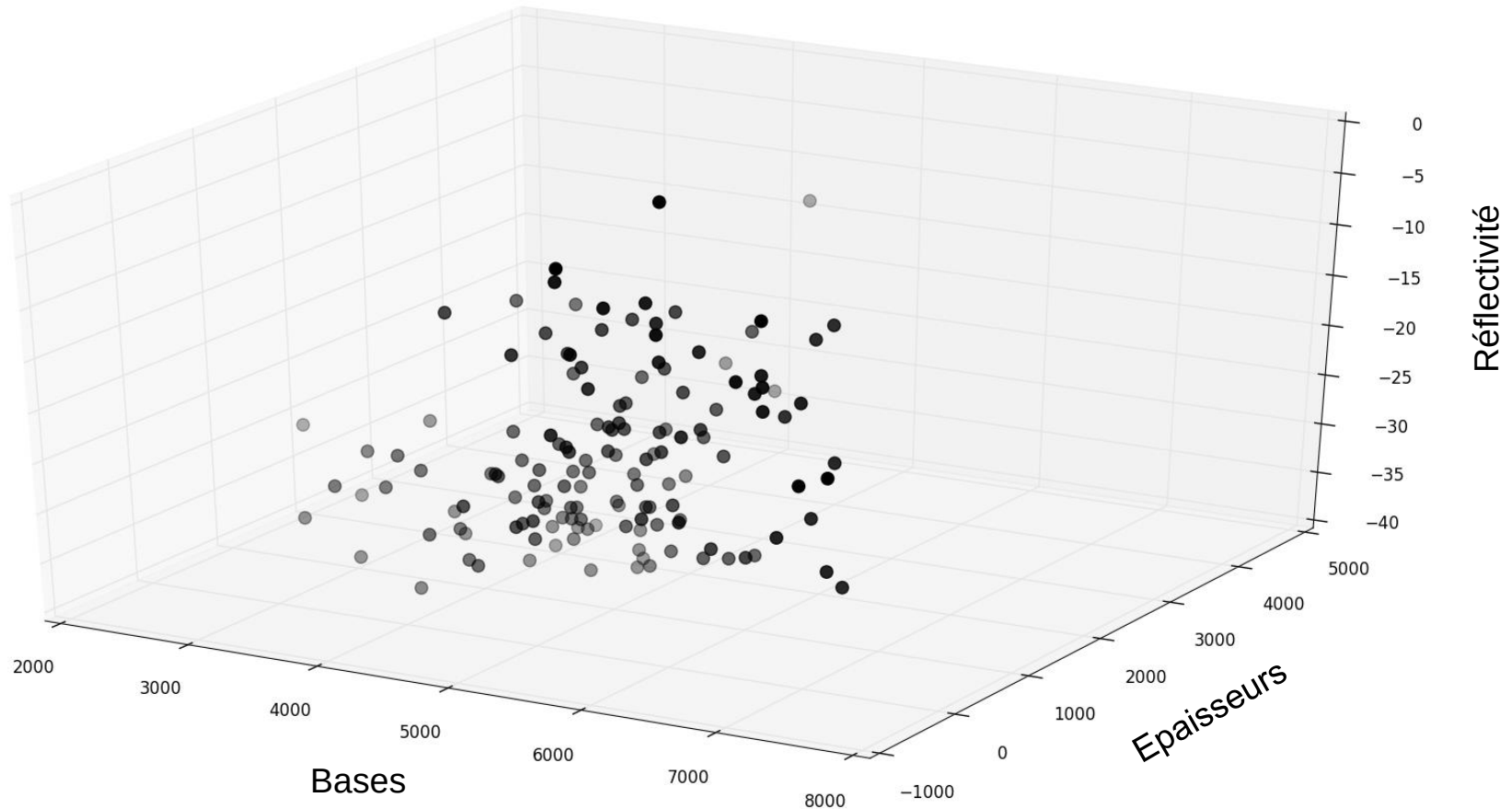


	Corrélation	P.value
Dim 1		
EcartTypeReflectivite	0,885	2,61E-49
Reflectivite	0,877	2,52E-47
Epaisseurs	0,832	2,43E-38
Durees	0,704	4,97E-23
Precipitation	0,687	1,44E-21
Sommets	0,561	2,11E-13
VitesseDoppler	-0,794	1,01E-32
Dim 2		
Bases	0,928	2,18E-63
Sommets	0,787	8,33E-32
Dim 3		
FractionLiquide	0,838	2,10E-39
Durees	0,410	3,06E-07
Dim 4		
HeureDeDebut	0,820	2,14E-36
HeureDeFin	0,712	1,05E-23

Variables
corrélées

Méthode de partitionnement K-Means

(pas JFM, durées > 30 min, détection radar $\geq 25\%$)

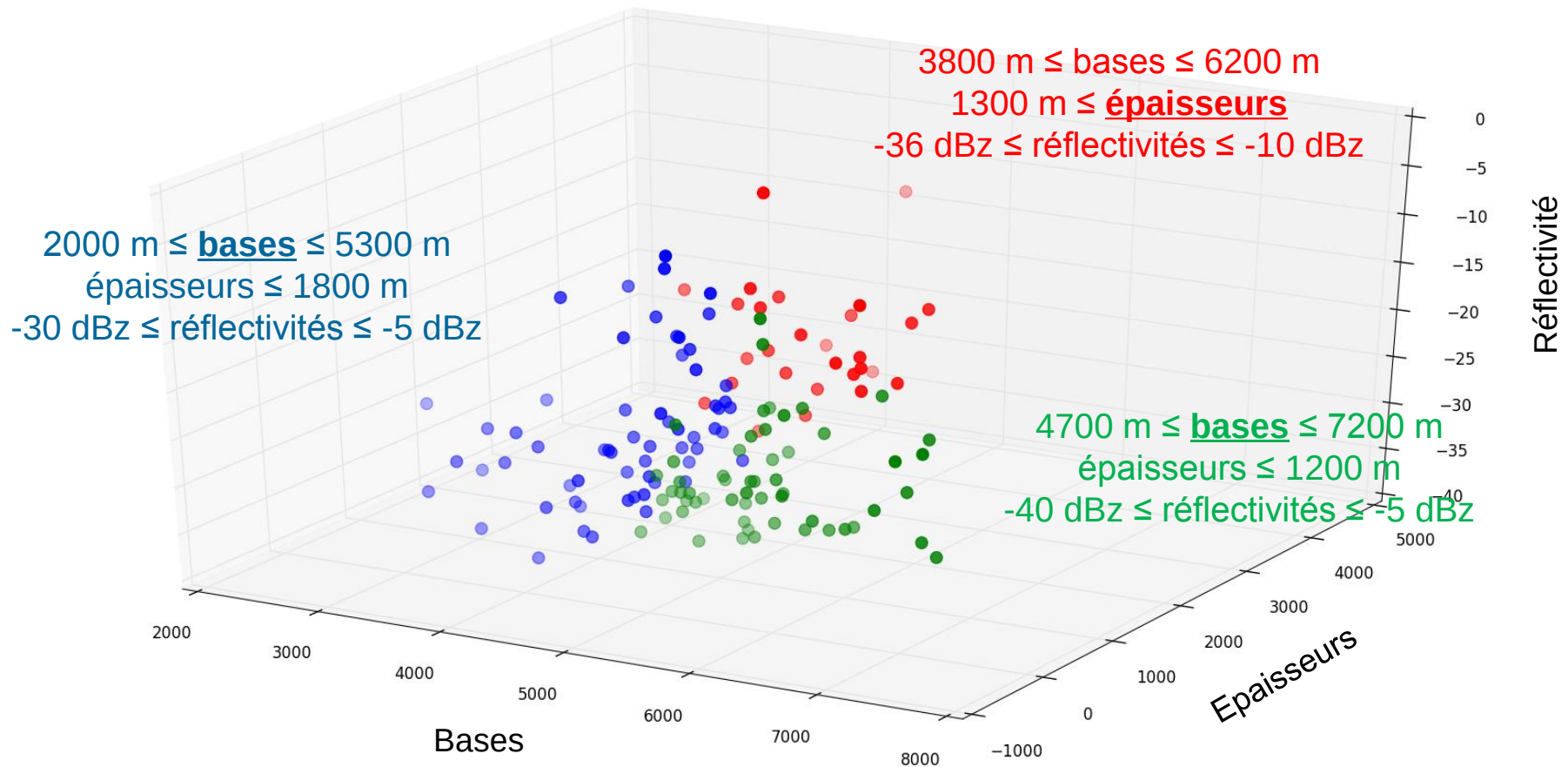


Utilisation des variables bases et épaisseurs de la base de données obtenue à partir des observations à Niamey en 2006 sur 145 nuages.

Méthode de partitionnement K-Means

Application

(pas JFM, durées > 30 min, détection radar $\geq 25\%$)

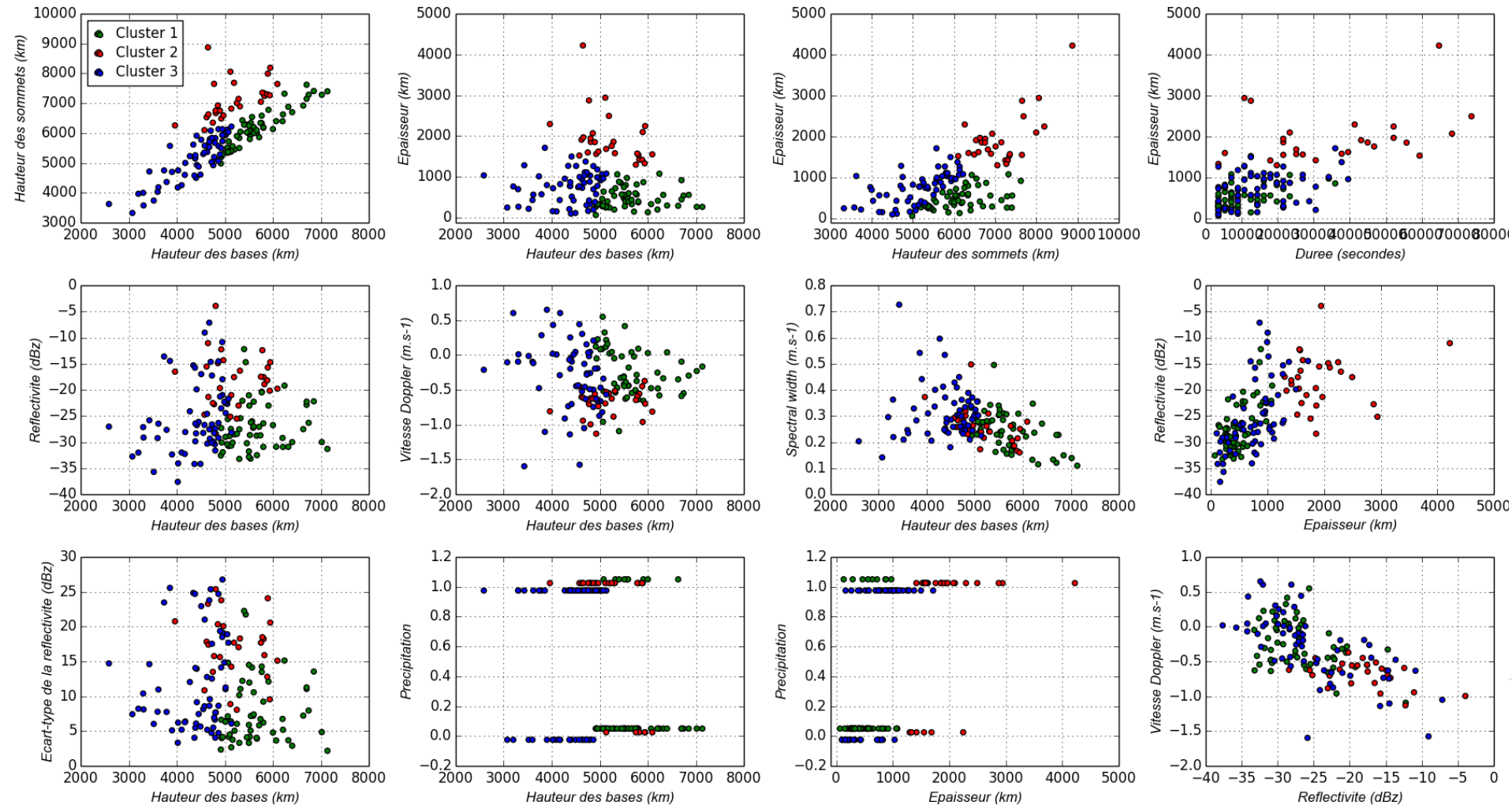


Utilisation des variables bases et épaisseurs de la base de données obtenue à partir des observations à Niamey en 2006 sur 145 nuages.

Méthode de partitionnement K-Means

Caractéristiques

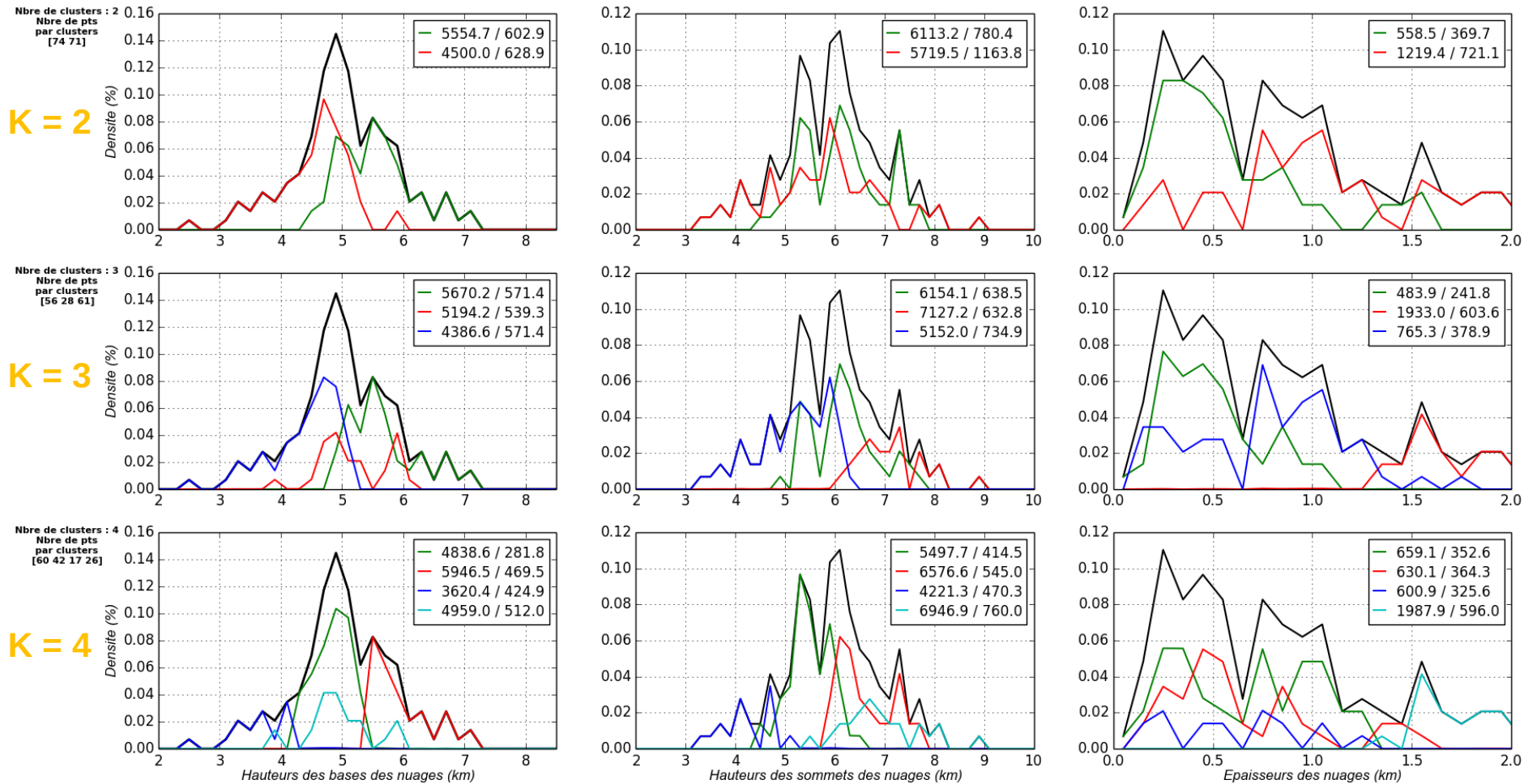
(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de partitionnement K-Means

Sensibilité à K (1)

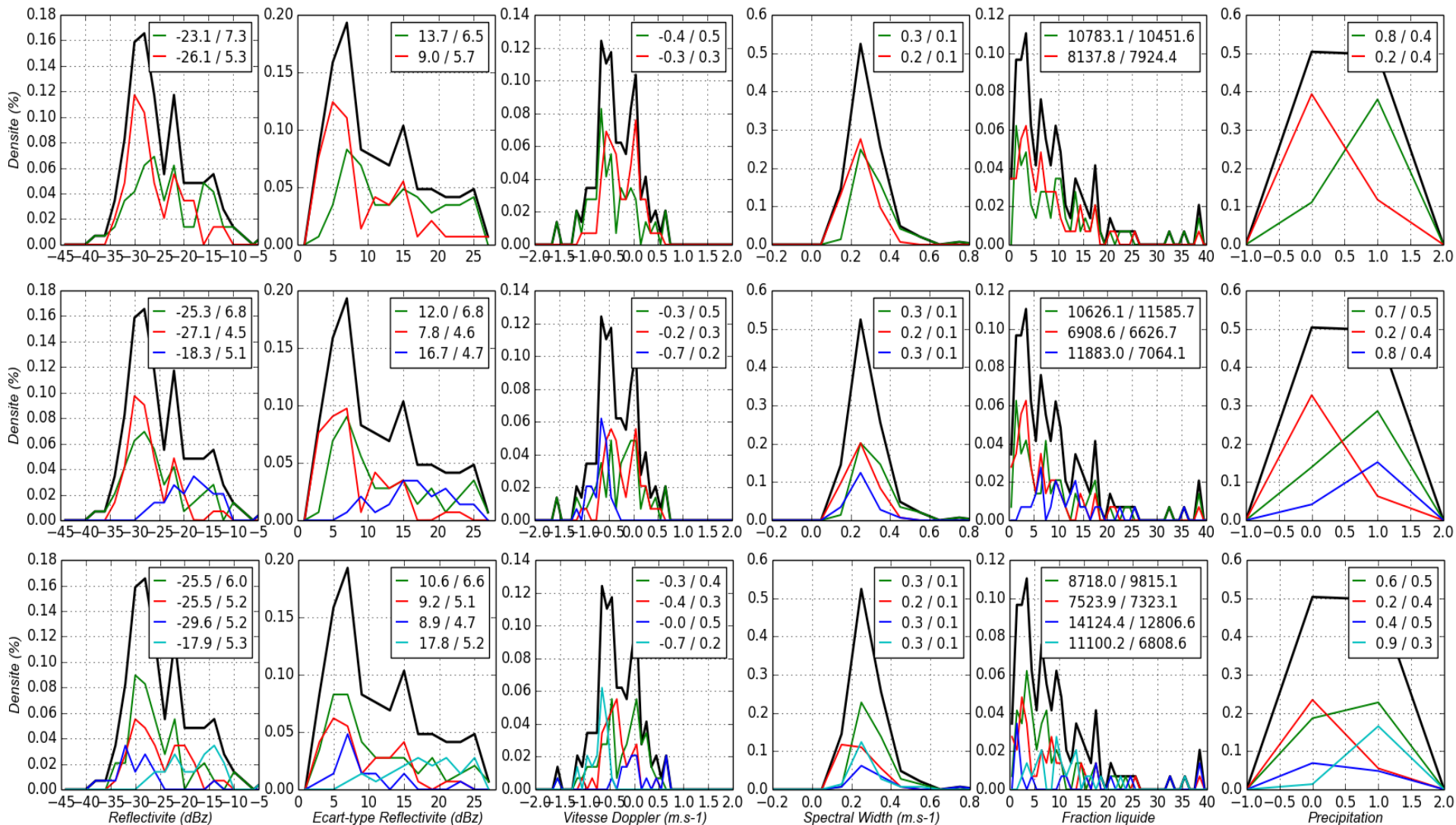
(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de partitionnement K-Means

Sensibilité à K (2)

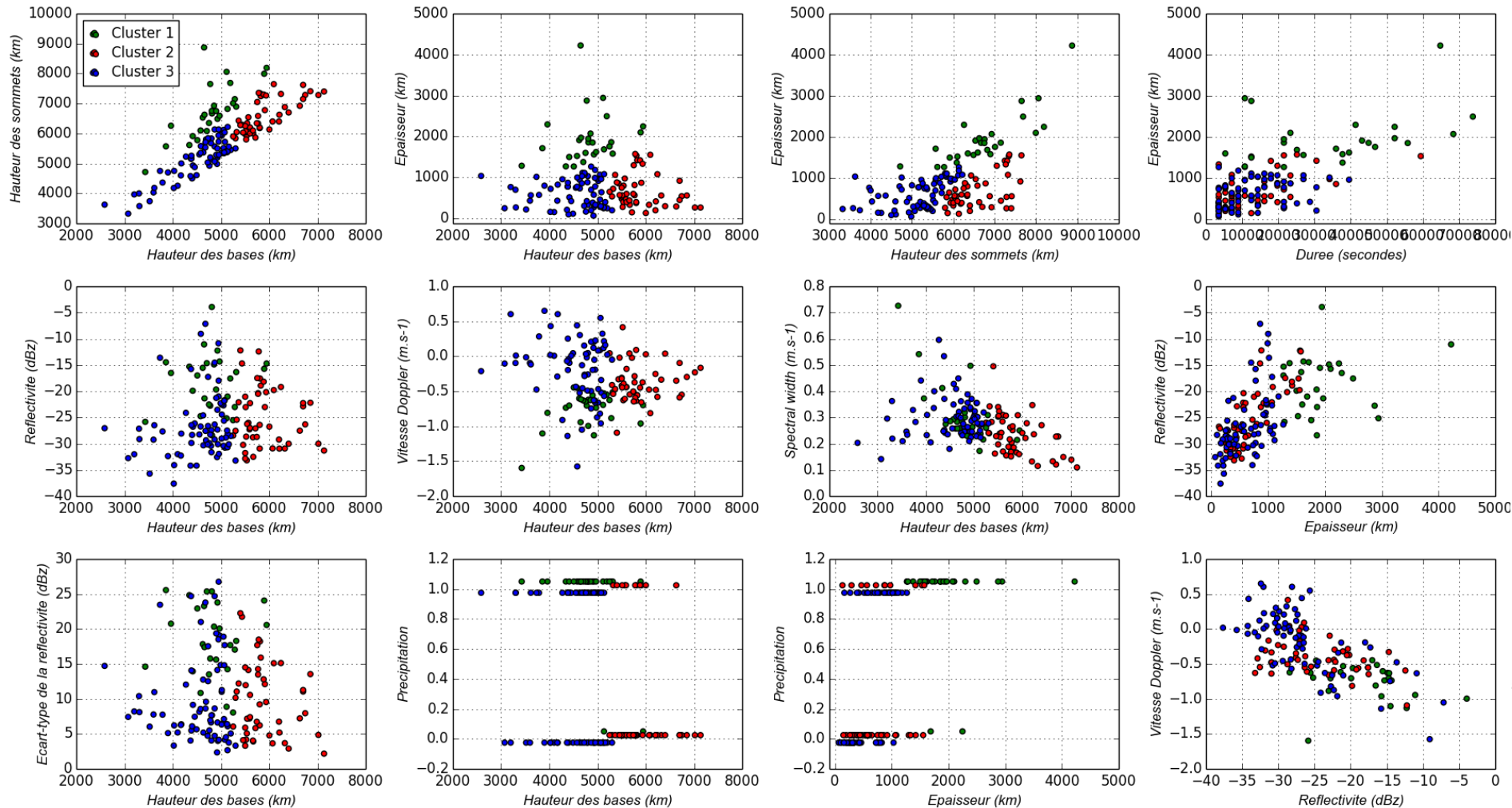
(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de hiérarchisation Ward

Caractéristiques

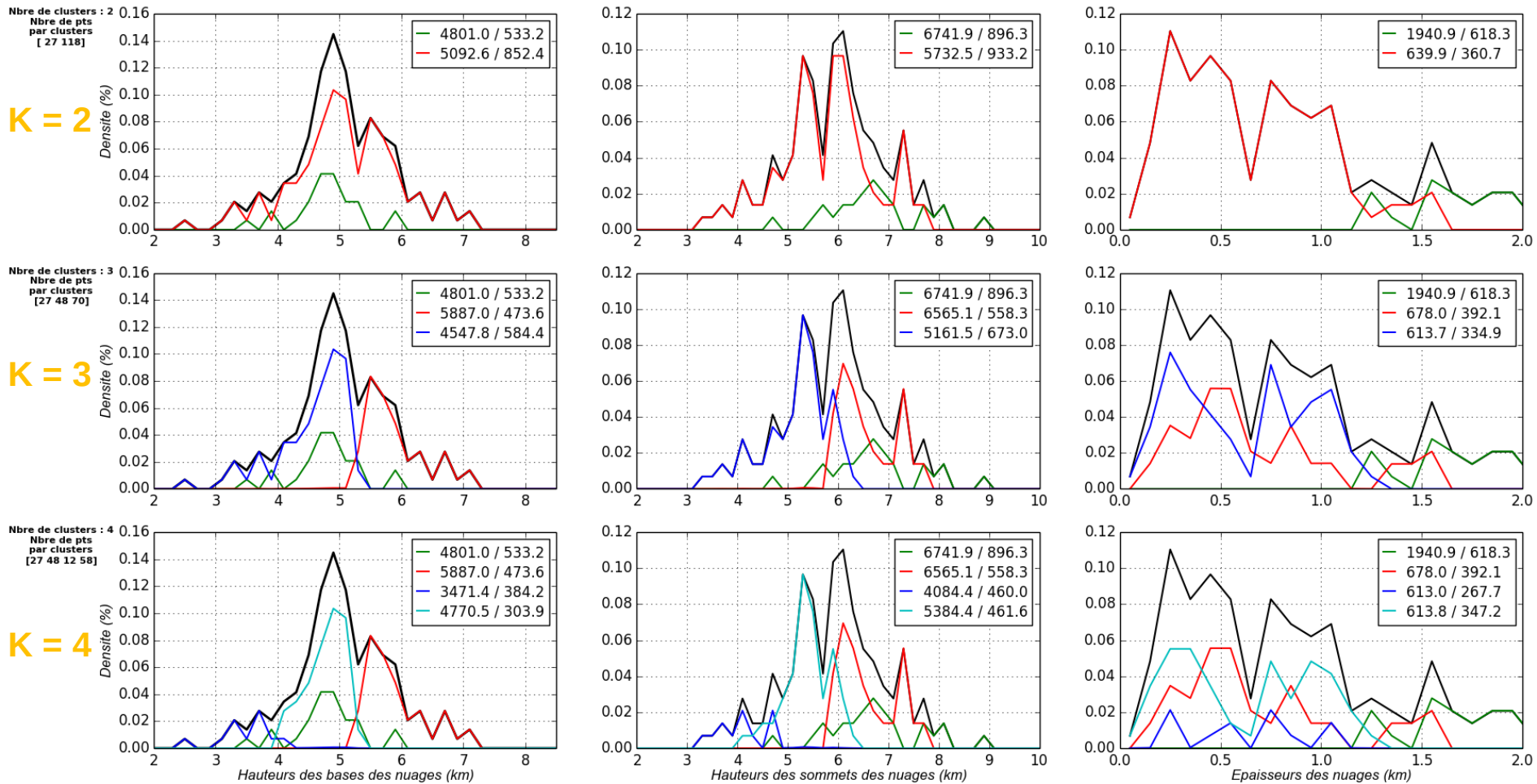
(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de hiérarchisation Ward

Sensibilité à K (1)

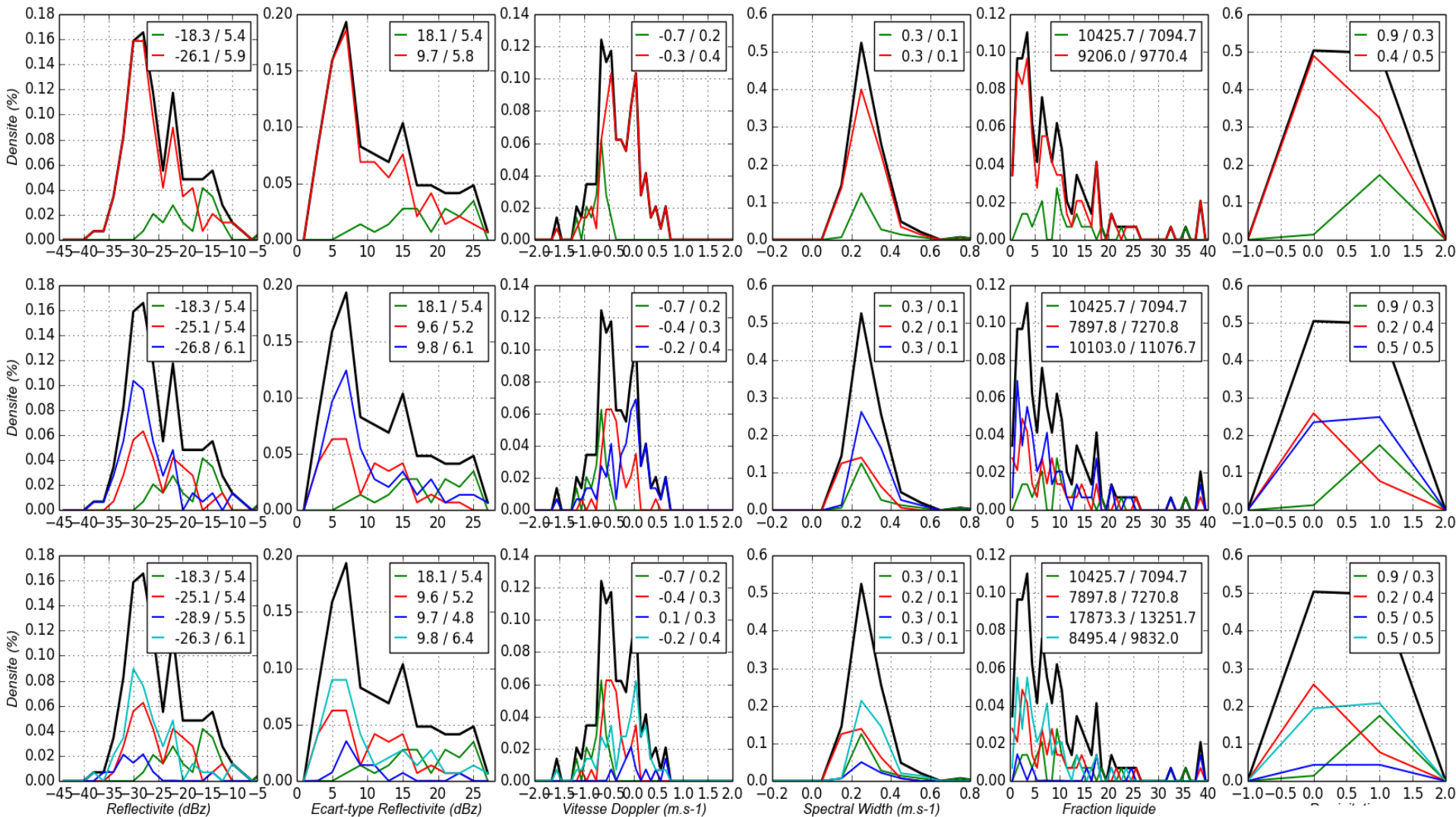
(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de hiérarchisation Ward

Sensibilité à K (2)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



Méthode de hiérarchisation Ward

Synthèse sur différents tests (1)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))

Variables d'entrées	Nombres de clusters différents
Bases / Épaisseurs	3
Bases / Épaisseurs / Réflectivités	2
Bases / Sommets / Réflectivités	2
Bases / Épaisseurs / Fraction Liquide / Heure de début	5

Nombres de clusters différents dans les dendrogrammes

D'après la méthode Elbow, K = 3
→ tests avec K = 3 mais aussi avec le nombre de clusters indiqué par le dendrogramme

Variables d'entrées	K = 3			K = 2		K = 5				
	C. 1	C. 2	C. 3	C. 1	C. 2	C. 1	C. 2	C. 3	C. 4	C. 5
Bases / Épaisseurs	27	48	70							
Bases / Épaisseurs / Réflectivités	60	17	68	60	85					
Bases / Sommets / Réflectivités	58	32	55	58	87					
Bases / Épaisseurs / Fraction Liquide / Heure de début	9	87	49			9	41	46	21	28

Nombres de nuages par clusters

Pour K = 3

Variables d'entrées	K = 3																	
	Cluster 1						Cluster 2						Cluster 3					
	Bases (m) mean	Sommets (m) mean	Épaisseurs (m) mean	Temps (h) mean	Heure de début mean	Réflectivité (dBZ) mean	Bases (m) mean	Sommets (m) mean	Épaisseurs (m) mean	Temps (h) mean	Heure de début mean	Réflectivité (dBZ) mean	Bases (m) mean	Sommets (m) mean	Épaisseurs (m) mean	Temps (h) mean	Heure de début mean	Réflectivité (dBZ) mean
Bases / Épaisseurs	moyens (+ bas)	hauts	épais	longs	matin (10h)	moyens (+ fortes)	hauts	hauts	fins	courts	matin (10h)	fortes	bas	bas	fins	courts	matin (10h)	fortes
Bases / Épaisseurs / Réflectivités	moyens (+ hauts)	hauts	épais	longs	matin (11h)	moyens (+ fortes)	bas	bas	fins	courts	matin (8h)	fortes	hauts	moyens (+ hauts)	fins	courts	matin (10h)	fortes
Bases / Sommets / Réflectivités	bas	bas	fins	courts	matin (9h)	fortes	hauts	hauts	fins	courts	matin (10h)	fortes	moyens (+ hauts)	hauts	épais	longs	matin (11h)	moyens (+ fortes)
Bases / Épaisseurs / Fraction Liquide / Heure de début	bas	bas	moyens (+ fins)	moyens	après-midi (14h)	fortes	moyens	moyens (+ hauts)	fins	courts	matin (8h)	fortes	moyens (+ hauts)	hauts	épais	longs	après-midi (14h)	moyens (+ fortes)

Pour les trois premiers tests, **trois types de nuages différents** : cirrus spissatus entraînés par des cumulonimbus, altocumulus/altostratus bas et altocumulus/altostratus hauts. En revanche, les regroupements du dernier test ne sont pas satisfaisants.

Méthode de hiérarchisation Ward

Synthèse sur différents tests (2)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))

Pour K = 2

Variables d'entrées	K = 2											
	Cluster 1						Cluster 2					
	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
Bases / Épaisseurs / Réflexivités	moyens (+ hauts)	hauts	épais	longs	matin (11h)	moyens (+ fortes)	moyens	moyens	fins	courts	matin (10h)	fortes
Bases / Sommets / Réflexivités	bas	bas	fins	courts	matin (9h)	fortes	hauts	hauts	moyens	moyens	matin (11h)	moyens (+ fortes)

Mélange des trois types de nuages vus précédemment.

Pour K = 5

Variables d'entrées	K = 5															
	Cluster 1								Cluster 2							
	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Fraction liquide	% nuages précipitant	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Fraction liquide	% nuages précipitant
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
Bases / Épaisseurs / Fraction Liquide / Heure de début	bas	bas	moyens (+ fins)	moyens	après-midi (14h)	fortes	fort	fort	hautes	hauts	fins	courts	matin (5h)	fortes	faible	faible

K = 5															
Cluster 3								Cluster 4							
Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Fraction liquide	% nuages précipitant	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Fraction liquide	% nuages précipitant
mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
basses	bas	fins	courts	matin (10h)	fortes	faible	moyen	moyens (+ hauts)	hauts	épais	longs	matin (5h)	moyens (+ fortes)	faible	fort

K = 5							
Cluster 5							
Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début	Réflexivité (dBz)	Fraction liquide	% nuages précipitant
mean	mean	mean	mean	mean	mean	mean	mean
hautes	hauts	moyens	moyens (+ longs)	soir (20h)	moyens (+ fortes)	moyen	moyen

Le cluster 5 possède des cirrus spissatus détrainés par des cumulonimbus et des altostratus / altocumulus hauts et quelques bas. L'heure de début est une variable d'entrée, elle rentre en compte dans le regroupement des nuages → pas forcément correct après analyse des quicklooks.

Synthèse sur différents tests (Ward – K=3 et K=2)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))

Variables d'entrées	K = 3																	
	Cluster 1						Cluster 2						Cluster 3					
	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
Bases / Epaisseurs	4801,0	6741,9	1940,9	9,4	10,05	-18,31	5887,0	6565,1	678,0	3,9	10,27	-25,07	4547,8	5161,5	613,7	3,6	10,69	-26,76
Bases / Epaisseurs / Réflexivités	5118,9	6536,6	1417,8	7,4	11,19	-18,50	3676,2	4188,2	511,9	3,1	8,05	-31,17	5307,7	5809,8	502,1	2,9	10,36	-28,39
Bases / Sommets / Réflexivités	4470,3	5027,9	557,6	3,1	9,66	-28,95	5942,6	6425,1	482,6	3,1	10,56	-28,39	5111,2	6568,0	1456,8	7,5	11,17	-17,88
Bases / Epaisseurs / Fraction Liquide / Heure de début	4127,6	4888,2	760,5	5,1	14,63	-24,54	4992,6	5577,2	584,6	3,2	8,05	-26,87	5286,6	6719,4	1432,8	7,6	13,90	-20,66

Variables d'entrées	K = 2											
	Cluster 1						Cluster 2					
	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
Bases / Epaisseurs / Réflexivités	5118,9	6536,6	1417,8	7,4	11,19	-18,50	4981,4	5485,5	504,0	2,9	9,89	-28,95
Bases / Sommets / Réflexivités	4470,3	5027,9	557,6	3,1	9,66	-28,95	5417,0	6515,4	1098,5	5,9	10,94	-21,75

Synthèse sur différents tests (Ward – K=5)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))

Variables d'entrées	K = 5															
	Cluster 1								Cluster 2							
	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Fraction liquide	% de nuages précipitants	Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Fraction liquide	% de nuages précipitants
	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean	mean
Bases / Epaisseurs / Fraction Liquide / Heure de début	4127,6	4888,2	760,5	5,1	14,63	-24,54	37449,2	77,8	5617,7	6105,6	487,9	3,1	5,66	-26,81	4753,2	24,4

Cluster 3							
Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Fraction liquide	% de nuages précipitants
mean	mean	mean	mean	mean	mean	mean	mean
4435,5	5106,3	670,8	3,3	10,17	-26,92	6955,4	47,8

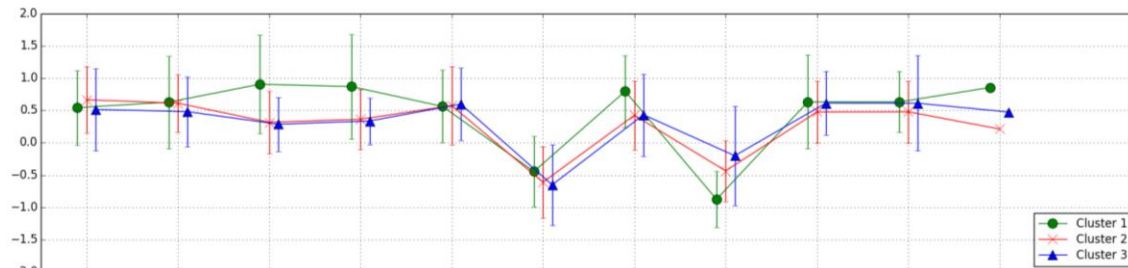
Cluster 4							
Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Fraction liquide	% de nuages précipitants
mean	mean	mean	mean	mean	mean	mean	mean
5249,4	7118,7	1869,2	8,9	5,02	-17,89	9585,5	81,0

Cluster 5							
Bases (m)	Sommets (m)	Épaisseurs (m)	Temps (h)	Heure de début (h)	Réflexivité (dBz)	Fraction liquide	% de nuages précipitants
mean	mean	mean	mean	mean	mean	mean	mean
5314,5	6419,9	1105,4	6,6	20,55	-22,74	11237,0	57,1

Synthèse sur différents test (Ward – $K=3$)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar $\geq 25\%$))

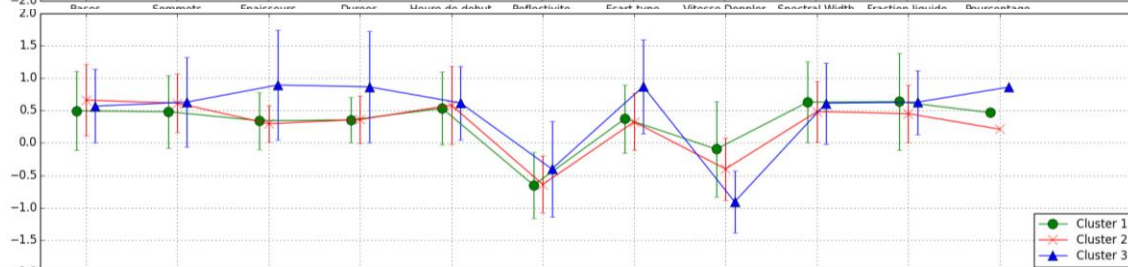
Bases /
Epaisseurs



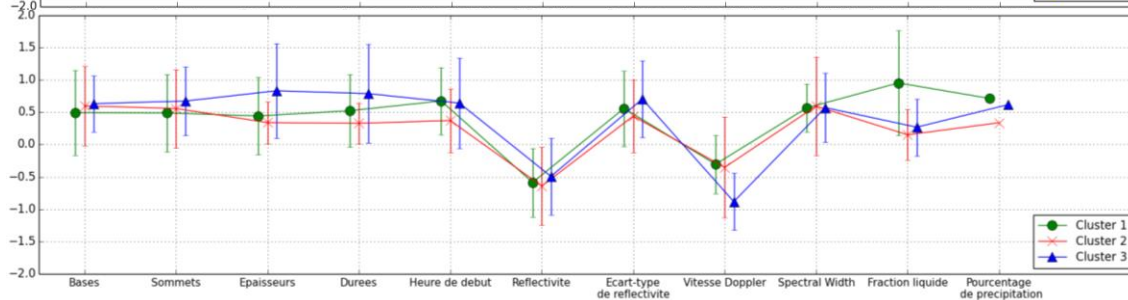
Bases /
Epaisseurs /
Réflectivités



Bases /
Sommets /
Réflectivités



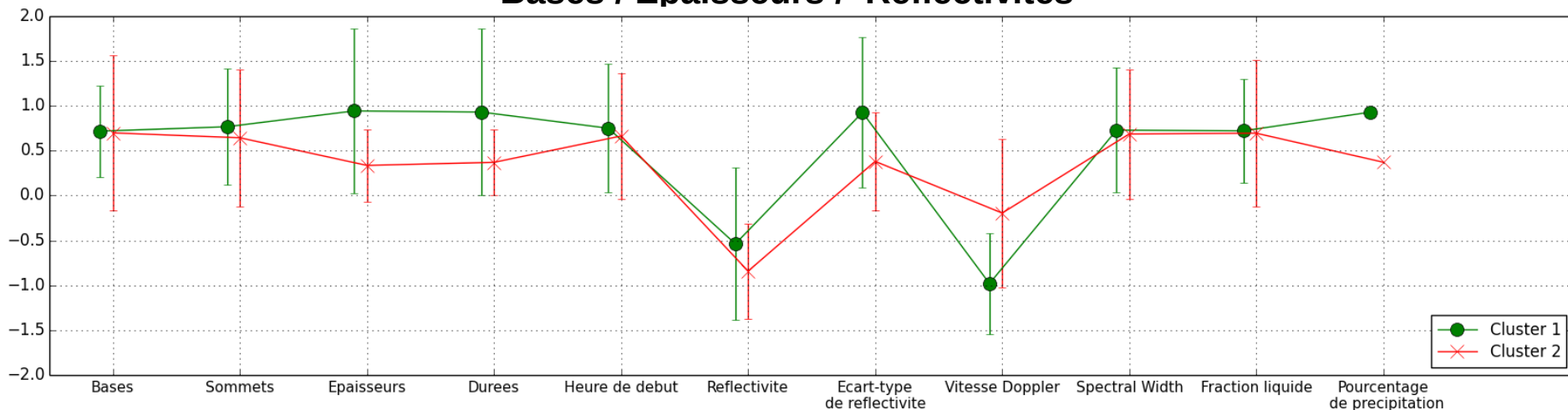
Bases /
Epaisseurs /
Fraction Liquide /
Heure début



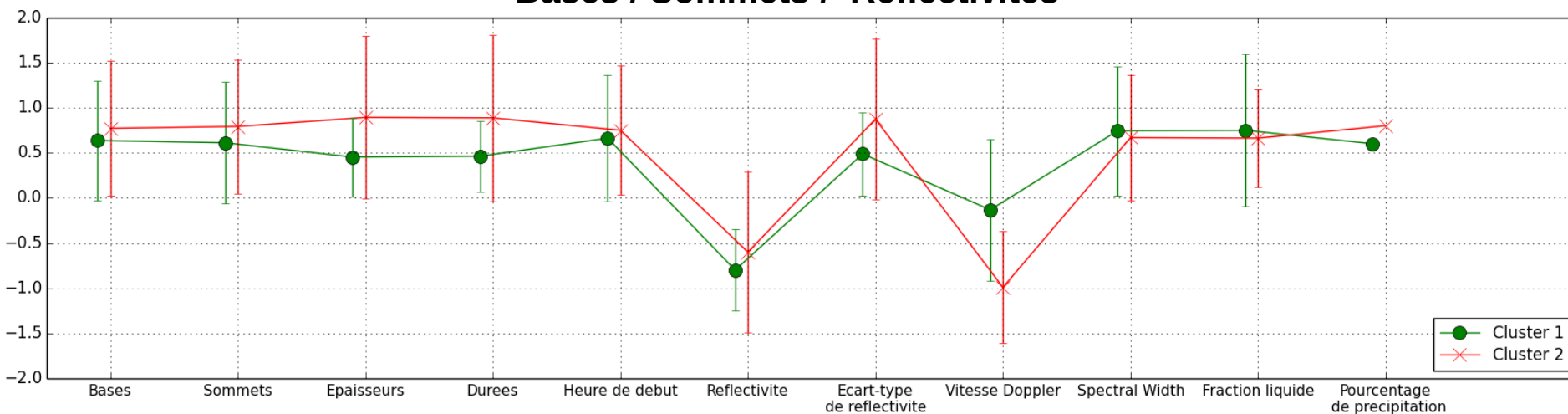
Synthèse sur différents test (Ward – $K=2$)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar $\geq 25\%$))

Bases / Epaisseurs / Réflectivités



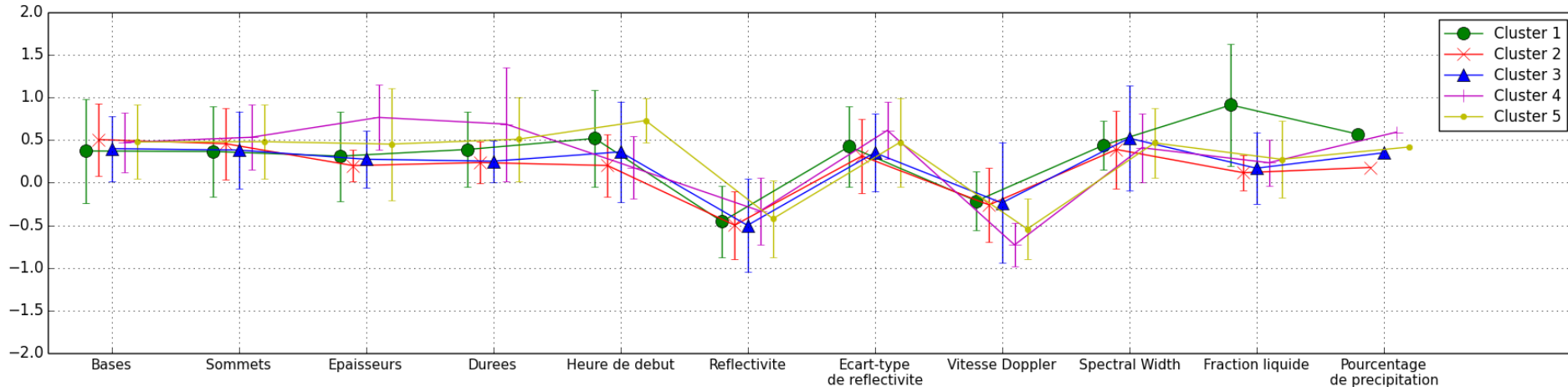
Bases / Sommets / Réflectivités



Synthèse sur différents test (Ward – $K=5$)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar $\geq 25\%$))

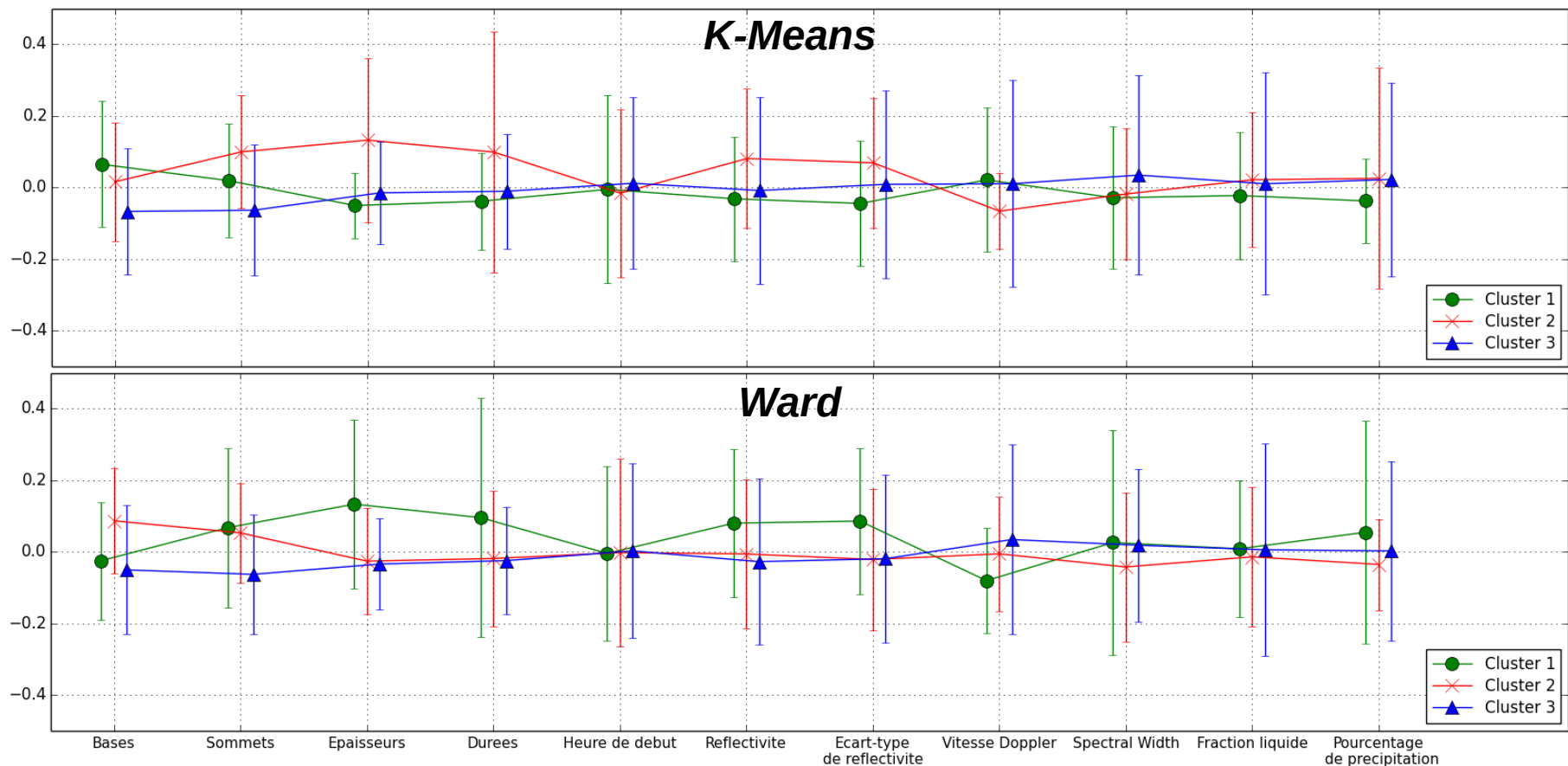
Bases / Epaisseurs / Fraction Liquide / Heure début



Les méthodes de clustering

Synthèse sur différentes méthodes avec Bases et Epaisseurs (K=3)

(Nuages représentatifs de la base de données (pas JFM, durées > 30 min, détection radar >= 25%))



K-Means	Ward	Types de nuages
1	2	Altostratus/Altostratus hauts
2	1	Cirrus Spissatus entraîné par un Cb
3	3	Altostratus/Altostratus bas

Trois types de nuages différents avec les deux méthodes mais différences de représentativités pour l'épaisseur, la durée, la réflectivité, la vitesse doppler et la spectral width. Davantage de difficulté à représenter la vitesse doppler.

Perspectives dans les mois prochains

La caractérisation de ces nuages et de leur environnement

- Après avoir **obtenu les différentes familles de nuages de mi-niveau** rencontrés à **Niamey** à partir du clustering,
 - **Réalisation des composites de profils thermodynamiques et de l'impact radiatif à la surface et au sommet de l'atmosphère** pour chaque famille,
 - **Documentation de l'environnement** pour chaque famille : circulation atmosphérique, positionnement par rapport à l'ITD, heat low, etc.
 - **Extrapolation de l'étude à l'Afrique de l'Ouest** en utilisant,
 - **D'autres jeux de données sols** (données FENNEC (Visite à Leeds, Angleterre) et AMMA),
 - **Des produits satellitaires,**
 - Analyse des caractéristiques de ces nuages (dans CloudSat/CALIPSO/AIRS),
 - **Validation des données par rapport aux caractéristiques 1D :**
Quels sont les jeux de données les plus appropriés pour mon étude ?
- Comparer tous les produits et les trier en fonction de leur pertinence
(enjeux méthodologiques de ce travail)

La caractérisation de ces nuages

Les jeux de données

Jeux de Données	Sols (*et aéroportées) AMMA (ARM Niamey 2006) FENNEC (Sahara)* AMMA - CATCH	Satellites										Analyses et ré-analyses ECMWF
		A-Train					Megha-tropiques		MetOp	Meteosat		
		CloudSat-CALIPSO	PARASOL	AQUA			SCARAB	SAPHIR	IASI 1 et 2	Meteosat	GERB	
Année		2006-2015	2003-2014	2002-2015	2002-2015	2002-2015	2011-2015		2006-2015	1981- -	2002- -	
Instruments	Lidar, radar et radiosondages	Lidar et radar	Radiomètre optique	Radiomètre	Sondeur infra-rouge	Imageur	Radiomètre	Sondeur micro-onde	Sondeur infra-rouge	Imageur	Radiomètre	
Profils verticaux de nébulosité (description couverture nuageuse)	X	X										
Nuages												
- Masques nuageux (Occurrence)	X	X			X	X	X		X	X		
- Classification nuageuse	X	X								X		
Fréquence temporelle (cycle diurne)	X		X			X				X		
Micro-physiques	X	X	X		X	X			X			
Profils thermo-dynamiques	X				X			X	X			X
Rayonnement												
- En surface	X	X	X	X						X		
- Au sommet		X	X	X			X				X	
Aérosols	X	X	X			X			X	X		

Perspectives pour la 2^{ème} année

L'identification des processus à l'origine de ces nuages

- **Analyse des conditions thermodynamiques** en établissant des schémas conceptuels indiquant les **processus à l'origine de ces nuages**,
- Simulation des différents types de nuages de mi-niveau (modèle 1D Més0-NH) en se concentrant sur certains paramètres :
 - Profils thermodynamiques,
 - Cycle diurne,
 - Position de l'isotherme 0°C (cf. Johnson et al., 1999),
 - Influence du vent,
 - Sensibilité aux profils d'aérosols,
 - Sensibilité à la résolution verticale.
- **Etude de l'impact de ces nuages sur la circulation de la mousson africaine** (en collaboration avec un post-doc - AMMA 2050)
 - Etudes diagnostics à partir du modèle Més0-NH en 2D afin de comprendre le rôle du nuage dans le bilan énergétique mais également les mécanismes à l'origine de la boucle de rétroaction nuage/rayonnement/dynamique.

Perspectives pour la 3^{ème} année

La représentation dans les modèles de ces nuages

- **Evaluation de la représentation des nuages de mi-niveau dans différents modèles**

Type	Résolution	Exemples
CRM (Cloud Resolving Model)	2 à 4 km	Runs CASCADE (4 km)
Analyses de centres de prévision	10 à 20 km	ARPEGE et ECMWF
Modèles de climat de CMIP5	100 à 300 km	ARPEGE et LMDZ

- La **représentation de ces nuages** de mi-niveau dans les modèles n'est **peut-être pas correcte** :
 - Soit parce que le modèle ne crée pas les bonnes conditions thermodynamiques nécessaires à la formation de ces nuages,
 - Soit parce que le modèle crée les bonnes conditions thermodynamiques nécessaires à la formation de ces nuages mais finalement il ne les forme pas.